

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

## Electrostatic interactions in the GCN4 leucine zipper: Substantial contributions arise from intramolecular interactions enhanced on binding

ZACHARY S. HENDSCH AND BRUCE TIDOR

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307

(RECEIVED December 31, 1998; ACCEPTED March 10, 1999)

### Abstract

The GCN4 leucine zipper is a peptide homodimer that has been the subject of a number of experimental and theoretical investigations into the determinants of affinity and specificity. Here, we utilize this model system to investigate electrostatic effects in protein binding using continuum calculations. A particularly novel feature of the computations made here is that they provide an interaction-by-interaction breakdown of the electrostatic contributions to the free energy of docking that includes changes in the interaction of each functional group with solvent and changes in interactions between all pairs of functional groups on binding. The results show that (1) electrostatic effects disfavor binding by roughly 15 kcal/mol due to desolvation effects that are incompletely compensated in the bound state, (2) while no groups strongly stabilize binding, the groups that are most destabilizing are charged and polar side chains at the interface that have been implicated in determining binding specificity, and (3) attractive intramolecular interactions (e.g., backbone hydrogen bonds) that are enhanced on binding due to reduced solvent screening in the bound state contribute significantly to affinity and are likely to be a general effect in other complexes. A comparison is made between the results obtained in an electrostatic analysis carried out computationally and simulated results corresponding to idealized data from a scanning mutagenesis experiment. It is shown that scanning experiments provide incomplete information on interactions and, if overinterpreted, tend to overestimate the energetic effect of individual side chains that make attractive interactions. Finally, a comparison is made between the results available from a continuum electrostatic model and from a simpler surface-area dependent solvation model. In this case, although the simpler model neglects certain interactions, on average it performs rather well.

**Keywords:** coiled coil; ion pairs; protein electrostatics; protein stability; salt bridges

Electrostatic interactions play a significant role in the structure and function of biological macromolecules. For example, hydrogen bonding between backbone polar groups is a fundamental feature of regular elements of protein secondary structure, molecular recognition of nucleic acid sequences frequently involves a binding interface rich in charged and polar groups, and enzyme catalysis often requires stabilization of charged intermediates in a reaction pathway. While the coulombic attraction between complementary polar and charged groups is clearly favorable, a substantial desolvation penalty must concomitantly be incurred to form such interactions upon protein folding or binding. In a recent study, we used continuum electrostatic calculations to examine 21 salt bridges in nine proteins to assess their contribution to protein folding (Hendsch & Tidor, 1994). Surprisingly, the electrostatic contribution of most of these salt bridges was found not to be stabilizing, and, in

fact, appeared to be destabilizing. That is, the electrostatic desolvation penalty due to burying the charged side chains upon protein folding was generally not fully recovered in favorable electrostatic interactions in the folded state, and the effect was the largest for the most buried salt bridges studied. One suggestion from the work is that the replacement of salt bridges with hydrophobic groups of similar size and shape could lead to more stable proteins. A further suggestion advanced is that compensated electrostatic interactions, even if destabilizing, could enhance specificity by dramatically disfavoring arrangements in which polar and charged groups are buried but not compensated (Hendsch & Tidor, 1994; Sindelar et al., 1998).

In an experimental study involving combinatorial mutagenesis of a salt bridge triad in Arc repressor, Waldburger et al. (1995) found that simultaneous hydrophobic substitutions for all three members of the salt bridge triad produced stability enhancements in the range of 1–2½ kcal/mol per monomer. In other work, using a peptide model system that partitioned between aqueous and organic phases representing the unfolded and folded states of pro-

Reprint requests to: Bruce Tidor, Department of Chemistry, Room 6-135, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307; e-mail: tidor@mit.edu.

teins, respectively, Wimley et al. (1996) estimated the electrostatic contribution of a salt bridge to protein folding to be around +4 kcal/mol. Thus, there is experimental support for the notion that salt bridge formation may be electrostatically destabilizing for protein folding.

Theoretical studies have led to similar conclusions about neutral hydrogen bonding groups and their electrostatic effect on folding. Using finite-difference Poisson-Boltzmann calculations, Honig and coworkers (Yang & Honig, 1995a, 1995b; Yang et al., 1996) showed that the electrostatic contribution of backbone hydrogen bonds to  $\alpha$ -helix,  $\beta$ -sheet, and turn formation is unfavorable. Using free energy simulation methods, Wang et al. (1996) showed that the electrostatic contribution of hydrogen bonds to  $\alpha$ -helix formation is unfavorable.

Here we examine further the role of electrostatics on protein thermodynamics, with a focus on protein binding rather than folding. Because charged side chains on the surfaces of folded proteins are likely to be less solvated than in the unfolded state, one might anticipate a smaller desolvation penalty for binding than folding, perhaps making these salt bridges electrostatically stabilizing. Moreover, rather than examine individual pairwise interactions, here we study each polar or charged functional group and its interactions with solvent and all other polar or charged groups in the complex. This analysis highlights not only the desolvation penalty and direct electrostatic interactions formed intermolecularly in the complex, but also shows fairly substantial energetics arising from "indirect" intramolecular electrostatic interactions (i.e., those resulting from reduced solvent screening of intramolecular interactions in the bound state compared to the unbound).

The computational analysis was carried out for the rigid docking in a model system, the two individual helices forming the GCN4 leucine zipper. This system was chosen because it has been extensively studied by a wide variety of experimental techniques. Crystal structures are available for both the leucine zipper dimerization motif alone (O'Shea et al., 1991) and for the leucine zipper and adjacent basic nucleic-acid-binding domain in complex with DNA (Ellenberger et al., 1992). Moreover, its small size makes it tractable for continuum electrostatic calculations. The model of the GCN4 leucine zipper used here is the 1.8 Å resolution crystal structure of O'Shea et al. (1991), which is derived from two 33-residue peptides with an N-terminal acetyl capping group. Each peptide adopts a right-handed  $\alpha$ -helical conformation, and the pair forms a parallel coiled-coil homodimer with left-handed superhelical twist. Leucine zipper sequences have a heptad repeat and are well represented by a helical wheel showing the sequence arranged into the seven distinct positions around the helix, with each position represented by the letters a–g (see Fig. 1).

The GCN4 leucine zipper structure (O'Shea et al., 1991; Ellenberger et al., 1992) forms a parallel coiled coil with local twofold symmetry throughout most of its length; there is evidence for slight fraying at the ends of the helices. The dimer interface is almost entirely hydrophobic, being made up of the conserved leucines (at position d) and principally valines (at position a; the exception is asparagine at position 16), which pack in a "knobs-into-holes" fashion (Crick, 1953). In the leucine layers, leucine residues pack against each other around the pseudo twofold axis as in a handshake and also pack against the following residue in the sequence (position e). In the alternate layers, valines (and asparagine at position 16) pack around the twofold axis and against the previous position (position g). This packing arrangement effectively sequesters the hydrophobic side chains at positions a and d

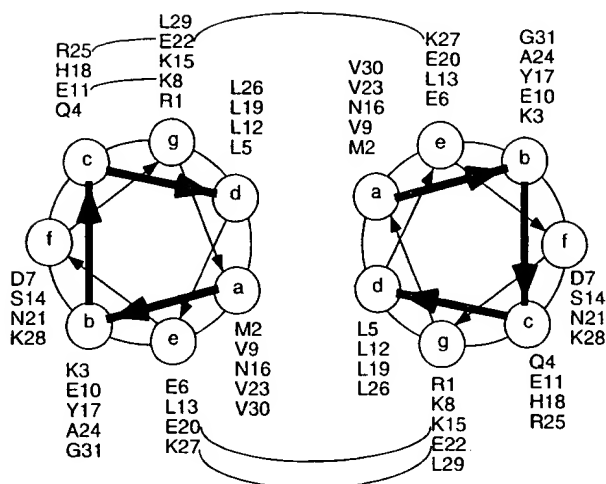


Fig. 1. Sequence of the GCN4 leucine zipper represented as a helical wheel. Salt bridges present crystallographically (O'Shea et al., 1991; Ellenberger et al., 1992) are indicated.

away from solvent. The buried asparagine at position 16 hydrogen bonds to Asn16 in the other monomer. Five salt bridges are found in both crystal structures. Three of these (Glu20A–Lys15B, Lys27A–Glu22B, and Lys27B–Glu22A) are interhelical salt bridges between side chains at the e and g positions. The other two (Lys8A–Glu11A and Glu22A–Arg25A) are intrahelical.

The leucine zipper family has proved to be a fruitful system for studying specificity. Mutation of Asn16 of GCN4 has resulted in a coiled coil that forms a mixture of dimers and trimers when this asparagine is mutated to valine (Harbury et al., 1993; Potekhin et al., 1994), alanine (Gonzalez et al., 1996b), aminobutyric acid (Gonzalez et al., 1996a), glutamine (Gonzalez et al., 1996c), or norleucine (Gonzalez et al., 1996c). Mutations that introduce asparagines at other a positions in GCN4 have been found to affect whether homodimers or heterodimers are formed (Zeng et al., 1997) and asparagine residues in the a position of other coiled coils have been found to affect the orientation (parallel vs. antiparallel; Lumb & Kim, 1995). Transplanting the e and g position residues, from Jun and Fos into the GCN4 background results in a pair of peptides that reproduce the heterodimeric behavior of Jun and Fos (O'Shea et al., 1992), and much of this specificity has been found to be due to two glutamic acid residues from Fos and two lysine residues from Jun (John et al., 1994), suggesting that electrostatic interactions are important for this specificity. Changes to the oligomerization state of GCN4 may be effected by altering the size and shape of hydrophobic residues at the a and d positions as well, with the appropriate changes resulting in preferred dimer, trimer, or tetramer states (Harbury et al., 1993).

## Results

The total electrostatic contribution to complex formation of the GCN4 leucine zipper structure was calculated using continuum electrostatics. In this study, complex formation was defined as the rigid docking of the two preformed helices from infinite separation to the crystal structure conformation. The total electrostatic contribution was calculated to be *destabilizing* by 15.0 kcal/mol.

To understand further the energetics of binding, the total electrostatic binding free energy was decomposed into a sum of terms representing individual pairwise and solvation interactions. The complex was divided into groups, with a separate group formed for each backbone amino group (H-N-C $\alpha$ ), backbone carbonyl (C=O), and side chain. For each group, three types of terms were calculated: a solvation term, direct terms, and indirect terms. The solvation term is the loss of solvent interactions of a group upon complex formation (such as by becoming buried). A direct term is an intermolecular solvent-screened coulombic interaction between two groups in the bound state of the complex. For example, the electrostatic attraction between two members of a salt bridge formed across the dimer interface would be measured as a direct term. An indirect term is an intramolecular interaction between two groups on the same component of the complex (i.e., helix here) and is the difference in the solvent-screened coulombic interaction between the groups in the bound and unbound state. Since each helix was held in the same conformation in the bound and unbound state, the coulombic portion of this term cancels, leaving only the difference in solvent screening (which is generally less effective in the bound state). An example of this would be an  $i$  to  $i + 4$   $\alpha$ -helical hydrogen bond that was buried at the dimer interface. The indirect interaction between the two members of this hydrogen bond would favor complex formation, since solvent would screen their favorable electrostatic interaction less in the bound state than in the unbound state. An important result of the work reported here is that energetics due to indirect effects are significant, though the importance of this term has generally not been recognized (Elcock & McCammon, 1996; Oberoi et al., 1996; Chong et al., 1998; Kangas & Tidor, 1998).

Summing each type of term for every group in the GCN4 leucine zipper (in a manner that counts each interaction once) showed that the electrostatic effect of complex formation is destabilizing because the unfavorable solvation term (24.0 kcal/mol) is only partially compensated by the favorable direct (−4.6 kcal/mol) and indirect (−4.4 kcal/mol) terms (Table 1). This extends our earlier observation that salt bridges only rarely provide electrostatic stabilization for protein folding due to underrecovery of desolvation energy (Hendsch & Tidor, 1994). The net effect of electrostatics on docking this peptide dimer was found to be destabilizing for essentially same reason. What is surprising here is that the direct term, representing intermolecular interactions, is relatively small and of the same size as the indirect term, representing enhanced intramolecular interactions.

**Table 1.** Contributions to  $\Delta\Delta G^{elec}$  by group and interaction type<sup>a</sup>

Interaction	Side chain	Backbone	Total <sup>b</sup>
$\Delta\Delta G^{solv}$	19.5	4.5	24.0 (1.0)
$\Delta\Delta G^{dir}$	−4.8	0.1	−4.6 (0.2)
$\Delta\Delta G^{indir}$	−1.5	−2.8	−4.4 (0.2)
Total	12.2	1.7	15.0 (1.0)

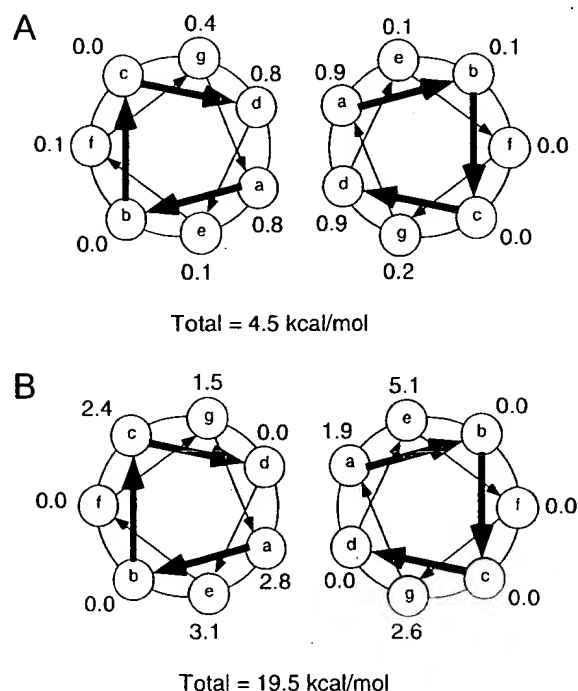
<sup>a</sup> All free energy values are in kcal/mol. Direct and indirect interactions that involve a backbone and a side chain group are divided equally between the two groups.

<sup>b</sup> The number in parenthesis reflects the calculational uncertainty. It reflects the error due to the grid representation of the molecule but does not reflect systematic error from the continuum model.

### Solvation terms

The desolvation energy of 24.0 kcal/mol is dominated by side-chain contributions (19.5 kcal/mol), with only 4.5 kcal/mol due to backbone groups. Groups along the dimer interface generally contribute most to the desolvation energy, as they are directly buried by complex formation. For example, the side chain of Asn16A (a position), which is almost completely buried upon docking, has a desolvation penalty of 2.8 kcal/mol, whereas the exposed side chain of Asn21A (f position), which loses no accessible surface area upon docking, has a computed desolvation penalty of 0.0 kcal/mol.

Since the position of a residue in the heptad repeat provides an approximate description of its position along the binding interface, the desolvation contribution was analyzed at the different positions (a–g) in the heptad repeat. It was expected that the desolvation penalty at each helical position would correlate with its burial, so the desolvation penalty at positions a and d would be greater than at e and g, which would be greater than b and c, which would be greater than f. This trend is followed essentially exactly for backbone groups (Figure 2A). The a and d positions contribute roughly four times what the e and g positions contribute. The backbone of the b, c, and f positions all contribute roughly equivalent, very small amounts to the total desolvation penalty. This trend is not followed for side chains, however, because the nature of the groups differs at each position; the a and d positions contain mostly hydrophobic side chains (all but Asn16), which incur no electrostatic desolvation penalty, so the e and g positions have a larger desolvation penalty than the a and d positions (Fig. 2). The b, c, and f positions, again, have very small desolvation penalties, except for position c in helix A, for which the value is 2.4 kcal/mol. All but



**Fig. 2.** Solvation penalty in kcal/mol summed over each heptad position for (A) backbone and (B) side-chain groups.



0.03 kcal/mol of this is due to the desolvation of Arg25A, which extends toward the interface to make an  $i, i + 3$  intrahelical salt bridge with the side chain of Glu22A (at position g) and becomes somewhat buried on docking. Therefore, in detail, the desolvation penalty of side chains in the leucine zipper depends on the composition, the position in the helical wheel, and the conformation of the side chain.

The placement of hydrophobic amino acids at the a and d positions reduces the desolvation cost for complex formation. If one assumed that the ratio of 4:1 that was seen for the backbone groups were to hold for side chains, then replacing the a and d side chains with those found in the e and g positions would cost 44 kcal/mol.

#### Interactions between individual pairs of chemical groups

Interactions between individual pairs of chemical groups can be separated into two distinct sets. Intermolecular (or direct) interactions are the solvent-screened coulombic interactions in the bound state between groups on different helices. Intramolecular (or indirect) interactions are the change in solvent screening upon complex formation for groups on the same helix. Interestingly, the net sum of all direct interactions (−4.6 kcal/mol) is very similar to the sum of all indirect interactions (−4.4 kcal/mol). These interactions have been decomposed further by the types of groups making the interactions. Side-chain–side-chain direct (−4.4 kcal/mol) and indirect (−1.5 kcal/mol) interactions and backbone–backbone indirect interactions (−2.8 kcal/mol) are large contributors and will be explored below.

#### Side-chain–side-chain interactions

Table 2 lists all intermolecular interactions greater in magnitude than 0.5 kcal/mol, all of which are side-chain–side-chain interactions. The Asn16A–Asn16B interaction is the strongest individual favorable interaction, which contradicts the popular notion that charge–charge interactions are usually stronger than polar–polar interactions. These two asparagines have the strongest interaction

because they are quite buried; the burial of these groups reduces solvent screening relative to the more exposed charge–charge interactions. It is important to note that the sum of the desolvation penalties for pairs of interacting groups is in all cases larger than the interaction between the groups. That is, the total electrostatic effect of each pairwise interaction is to destabilize the complex (see below).

Solvent exposed charge–charge pairings comprise the remainder of the significant intermolecular interactions. Four of these interactions are the attractive  $g \rightarrow e'$  interhelical interactions thought to be important for helix formation. Three of these interactions, between Glu20A–Lys15B, Glu22A–Lys27B, and Glu22B–Lys27A are each worth about −1.75 kcal/mol. The other  $g \rightarrow e'$  interaction, between Glu20B and Lys15A, is too distant to be called a salt bridge (6.2 Å at closest approach) and has a much weaker interaction (−0.6 kcal/mol). There is an interhelical repulsion between Glu20A at the e position and Glu22B at the g position that costs 0.5 kcal/mol. The symmetric interaction, between Glu20B and Glu22A, is only 0.3 kcal/mol.

The set of interactions among the side chains of residues Glu22A, Arg25A, and Lys27B consists of a Glu22A–Lys27B interhelical salt bridge (−1.8 kcal/mol), a Glu22A–Arg25A intrahelical salt bridge (−0.9 kcal/mol, due only to reduced solvent screening on binding), and an Arg25A–Lys27B interhelical repulsion (2.4 kcal/mol). This repulsion is a consequence of two positively charged side chains, each interacting with the same negative charge. Interestingly, it is present in both crystal structures (O'Shea et al., 1991; Ellenberger et al., 1992), but it is not known whether it exists in solution. Although the symmetry-related interaction is not made, it appears that it is involved in a crystal contact. Relaxing this repulsion by moving one of the positive charges away could be favorable electrostatically but would presumably also entail a hydrophobic cost (due to less burial of surface area and poorer packing; Nicholls et al., 1991).

#### Backbone–backbone interactions

Intramolecular backbone–backbone interactions (backbone dipolar groups within the same helix) contribute −2.8 kcal/mol to docking. These favorable interactions are due to reduced solvent screening upon complex formation, which enhances intrahelical hydrogen bonds by decreasing the effective dielectric constant for these interactions. Interestingly, this is offset somewhat by enhanced repulsion between backbone groups within the same turn of the helix. This is a simple result of the fact that parallel dipoles attract when arranged end to end and repel when aligned side by side.

Intermolecular backbone–backbone interactions (between the two helices), account for a net unfavorable total of 0.6 kcal/mol. Unfavorable interactions due largely to repulsions between parallel dipolar groups in the same turn of the two helices (such as the carbonyls and C $\alpha$ –N–H's related by pseudo-twofold symmetry) are partially offset by favorable interactions (−3.9 kcal/mol) between backbone dipolar groups on adjacent turns of the two helices. This is again a result of the fact that parallel dipoles will repel when arranged side by side but attract when placed end to end. It had been suggested early on that parallel coiled coils may be less stable than antiparallel ones due to "helix–dipole" effects (Landschulz et al., 1988). These results show that the repulsion is quite small because it is compensated rather remarkably by a set of attractions that are integral to coiled-coil structure. Moreover, the results also suggest that the two helix backbones do not interact as two large

**Table 2.** Strong interactions between individual groups<sup>a</sup>

Interaction <sup>b</sup>	Magnitude <sup>c</sup>
Asn16A–Asn16B	−2.1
Lys15A–Glu20B	−0.6
Lys15B–Glu20A	−1.8
Glu20A–Glu22B	0.5
Glu22A–Arg25A	−0.9
Glu22A–Lys27B	−1.8
Arg25A–Lys27B	2.4
Lys27A–Glu22B	−1.7

<sup>a</sup>All free energy values are in kcal/mol; all interactions of magnitude 0.5 kcal/mol or greater are included.

<sup>b</sup>Values for interhelical interactions are the solvent-screened coulombic interaction energy between the groups. The value for the intrahelical interaction (Glu22A–Arg25A) is the difference between their interaction in the bound and unbound state and reflects the change in screening by solvent of their interaction upon binding.

<sup>c</sup>Positive values indicate unfavorable interactions. The calculational uncertainty for all direct and indirect interactions is 0.2 kcal/mol. The largest uncertainty for a single interaction listed here is 0.02 kcal/mol for the Arg25A–Lys27B interaction.

macroipoles whose interactions are either favorable (antiparallel) or unfavorable (parallel), but instead the local interactions of individual backbone groups determine whether the backbone interactions are stabilizing or destabilizing (Hendsch & Tidor, 1994; Tidor, 1994; Prévost, 1996). In this case, these local interactions result in a substantial but incomplete cancellation of the predicted unfavorable interaction of the presumed repulsion.

#### Overall contribution of individual chemical groups

We define the total contribution of an individual group to docking by including its full desolvation penalty and half of its direct and indirect interactions (assigning the "other half" to its partner in each interaction; see Materials and methods). This contribution from each group is additive so the sum of the contribution for all groups equals the total electrostatic binding free energy. Note that a group's contribution in this sense is different from the effect on binding of mutating away its charge, since that would eliminate solvent interactions and all (not half) its interaction free energy with all other groups.

Table 3 lists the total electrostatic contributions for all polar and charged side chains at the e and g positions as well as for each Asn16 and Arg25. No other group has a contribution larger than 0.3 kcal/mol. All of the significant electrostatic contributions to complex formation come from side chains along the interface; side chains away from the interface and individual backbone groups contribute only a small amount (Fig. 3). All together, groups not included in Table 3 destabilize the complex by 1.8 kcal/mol (two times the standard error is 0.05 kcal/mol), so many small effects do add to a substantial amount. 1.7 kcal/mol of this 1.8 kcal/mol is due to contributions from the backbone, showing the overall contribution of the side chains away from the interface is negligible.

Interestingly, no chemical group makes a substantial stabilizing total electrostatic contribution to complex formation; the largest magnitude is  $-0.2$  kcal/mol from Glu22A. There are several groups that make large destabilizing contributions; the side chains of Asn16A, Asn16B, Glu20A, Glu20B, Glu22B, Lys27B, and Arg25A all have a contribution larger than 0.5 kcal/mol. It is interesting that all of these groups make strong attractive pairwise electrostatic interactions (Table 2) but still do not recover their desolvation penalties. However, all of these residues (except perhaps Arg25A) play a role in binding specificity (O'Shea et al., 1992; Harbury et al., 1993; Lumb & Kim, 1995; Nautiyal et al., 1995). This is consistent with the picture that electrostatics are generally destabilizing but contribute to specificity (Hendsch & Tidor, 1994; Sindelar et al., 1998). These results focus the hypothesis further by suggesting that the most destabilizing electrostatic interactions are frequently used to impart specificity.

The cost of mutating the charge for each group was computed as the group's desolvation penalty plus the sum of all of its intra- and intermolecular interactions. This is plotted against the total electrostatic contribution for each group (the desolvation energy plus one-half the sum of the interactions) in Figure 4. The contribution is systematically less favorable than the mutation cost, which indicates that most groups make attractive interactions with other charged and polar groups in the protein. Because of this, one would expect that using individual mutational energetics to measure the importance of a side chain to the binding free energy will overestimate a group's importance in approximately direct proportion to the strength of its interactions.

**Table 3.**  $\Delta\Delta G^{\text{contrib}}$  and  $\Delta\Delta G^{\text{mut}}$  of side chains by position<sup>a</sup>

Group	$\Delta\Delta G^{\text{contrib}}$ <sup>b</sup>	$\Delta\Delta G^{\text{mut}}$ <sup>c</sup>
<b>Position a</b>		
Asn16A	1.8	0.8
Asn16B	0.8	-0.3
<b>Position e,g</b>		
Arg1A	0.0	0.1
Glu6A	0.3	0.3
Lys8A	0.0	-0.2
Lys15A	0.1	-0.2
Glu20A	1.1	0.3
Glu22A	-0.2	-1.3
Lys27A	0.1	-0.5
Arg1B	0.0	0.0
Glu6B	0.3	0.2
Lys8B	-0.1	-0.2
Lys15B	-0.1	-0.7
Glu20B	0.9	0.8
Glu22B	1.3	0.6
Lys27B	3.8	4.0
<b>Position b</b>		
Arg25A	2.9	3.4
Arg25B	0.1	0.2

<sup>a</sup>All free energy values are in kcal/mol. Positive values represent unfavorable terms.

<sup>b</sup>The contribution of a group is the sum of its desolvation penalty and one-half of its interactions (see text). The largest calculational uncertainty for a single group is 0.5 kcal/mol for Arg25A.

<sup>c</sup>The mutation term is the sum of its desolvation penalty and its interactions. The largest calculational uncertainty for a single group is 0.5 kcal/mol for Arg25A.

The effect of pairwise mutations in which each of the three intermolecular salt bridges, the Asn16A-Asn16B interactions, and the distant ion pair (Glu20B-Lys15A) was individually mutated to its nonpolar isostere are presented in Table 4. The results show that none of these interactions stabilizes docking electrostatically. On average, each is destabilizing by about 2 kcal/mol.

#### Effect of parameterization on results

The error ranges reported above refer to computational uncertainty due to mapping the molecular boundaries and charge distributions onto a grid in the numerical calculation (see Materials and methods). Other sources of error include systematic error due to use of a relatively coarse grid, inaccuracies in the atomic radius and partial charge values used, uncertainty in the appropriate value for the internal dielectric, details of the definition of the dielectric boundary and potential weaknesses of the use of a continuum model, such as neglect of effects of electrostriction, dielectric saturation, solvent granularity, and effects of local hydrogen bonding (Rick & Berne, 1994; Marten et al., 1996). Here we investigate the significance of a number of these effects.

Trial calculations were carried out at grid spacings up to 4 grid units per Å. These showed only small variation in the total binding free energy (around 1 kcal/mol) as well as in the components. The

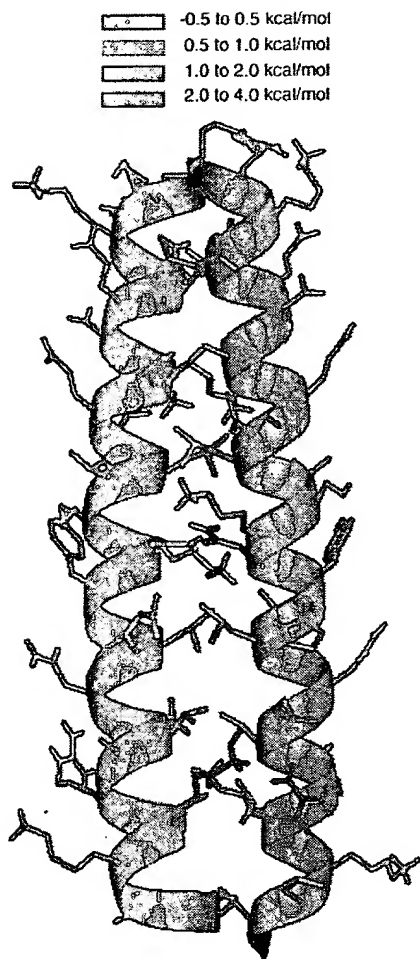


Fig. 3. Structure of the GCN4 leucine zipper color coded according to  $\Delta\Delta G^{\text{contrib}}$ . Most groups, in blue, essentially recover their desolvation penalty. Groups that do not and contribute strong destabilizing effects include a number of residues implicated in specificity determination through electrostatic patterning.

variation was under 0.1 kcal/mol for almost all of the groups. A few groups had larger variations, the largest being about 0.5 kcal/mol. Therefore, the coarse grid ( $65 \times 65 \times 65$ , about 1 grid unit per Å) provides an adequate representation of the leucine zipper complex and was used for the component analysis, since it requires substantially less computer time than the finer grids.

The dependence of the results on the internal dielectric was tested by repeating aspects of the calculation using a range of values for the internal dielectric constant from 1 to 10.3 (which is the dielectric constant of octanol; see Table 5). The electrostatic free energy of complex formation depends dramatically on the value used for the internal dielectric; the overall value as well as the  $\Delta\Delta G^{\text{hyd}}$  ( $= \Delta\Delta G^{\text{solv}} + \Delta\Delta G^{\text{indir}}$ ) and  $\Delta\Delta G^{\text{dir}}$  components scale roughly with the inverse of the internal dielectric, particularly for small values of the dielectric. This is in sharp contrast to the external dielectric, where  $\Delta\Delta G^{\text{elec}}$  ranges only from 12.8–16.3 kcal/mol as the external dielectric constant is varied from 40 to 200 ( $\Delta\Delta G^{\text{hyd}}$  remains roughly constant with the different values of

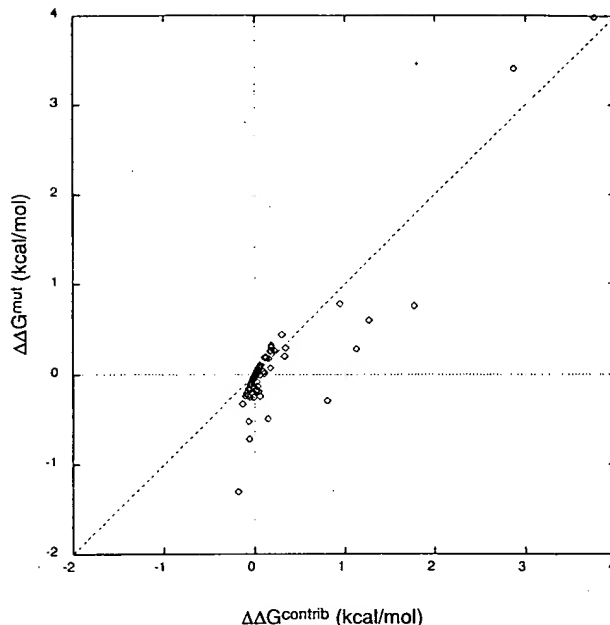


Fig. 4. Relationship of the additive electrostatic contribution of each group to binding ( $\Delta\Delta G^{\text{contrib}}$ ) to that which would be measured in a hypothetical experiment in which the polar and charged atoms of each group are individually neutralized ( $\Delta\Delta G^{\text{mut}}$ ).

the external dielectric and the change is due almost entirely to the solvent screening of  $\Delta\Delta G^{\text{dir}}$ ; results not shown). Moreover, even the “partitioning” of the overall electrostatic result into interactions arising from the same monomer ( $\Delta\Delta G^{\text{hyd}}$ ) vs. those arising from charges in the other monomer ( $\Delta\Delta G^{\text{dir}}$ ) depends somewhat on the value of the internal dielectric (e.g., the ratio changes from about 1:5.6 with  $\epsilon_{\text{int}} = 2$  to about 1:4.9 with  $\epsilon_{\text{int}} = 4$ ). However, this change is modest enough for the components that dominate the results for their relative magnitudes to be fairly independent of the value of the internal dielectric constant over the currently commonly used range of 2–4 (results not shown). All values of the internal dielectric constant tested predict that electrostatics desta-

Table 4. Pairwise mutation of bridging residues<sup>a</sup>

Bridging side chains	Pairwise mutation <sup>b</sup>	Desolvation penalty <sup>c</sup>	Bridging interaction <sup>d</sup>	All other interactions <sup>e</sup>
Asn16A–Asn16B	2.5	4.7	−2.1	−0.1
Glu20A–Lys15B	1.4	2.6	−1.8	0.6
Glu20B–Lys15A	1.2	1.5	−0.6	0.3
Glu22A–Lys27B	4.5	4.5	−1.8	1.8
Glu22B–Lys27A	1.8	2.7	−1.7	0.9

<sup>a</sup> All energy values are in kcal/mol.

<sup>b</sup> Total electrostatic effect of mutating to the pair of bridging side chains from hydrophobic isosteres.

<sup>c</sup> The sum of the desolvation penalties of the two side chains.

<sup>d</sup> The magnitude of the interaction between the two side chains.

<sup>e</sup> The sum of the interactions (direct and indirect) that the bridging side chains make with other groups in the complex.

**Table 5.** Dependence of  $\Delta\Delta G^{\text{elec}}$  and major components on the value of the internal dielectric constant<sup>a</sup>

$\epsilon_{\text{int}}$	$\Delta\Delta G^{\text{elec}}$	$\Delta\Delta G^{\text{dir}}$	$\Delta\Delta G^{\text{hyd}}$
1	66.6 (4.5)	-11.7 (0.7)	78.3 (4.6)
2	32.2 (2.1)	-7.0 (0.4)	39.2 (2.2)
3	20.7 (1.4)	-5.4 (0.2)	26.2 (1.4)
4	15.0 (1.0)	-4.6 (0.2)	19.6 (1.0)
6	9.3 (0.6)	-3.8 (0.1)	13.0 (0.6)
8	6.4 (0.4)	-3.3 (0.3)	9.7 (0.4)
10.3	4.5 (0.3)	-3.0 (0.1)	7.5 (0.3)

<sup>a</sup> All free energy values are in kcal/mol. Numbers in parenthesis indicate calculational uncertainty.

bilize complex formation. The results discussed in this paper were calculated using an internal dielectric constant of four, which, of the commonly used dielectric constants, predicts salt bridges to be the least destabilizing, so this may provide a lower bound on the destabilizing effects of charges (Hendsch & Tidor, 1994).

The effect of charge and radius parameters is shown in Table 6. Three charge and radius combinations were used: CHARMM, PARSE, and OPLS (for OPLS atomic radii were set to be  $2^{-5/6}\sigma$  except for hydrogen, for which a radius of 1.25 Å was used). Moreover, dependence on atomic radii alone was also studied using the CHARMM charge parameters with seven different radius parameters (see Table 6). This mixing of parameters could be somewhat inconsistent, but may provide a generous estimate of

**Table 6.** Dependence of  $\Delta\Delta G^{\text{elec}}$  and major components on the radius and charge parameters<sup>a</sup>

Parameters	$\Delta\Delta G^{\text{elec}}$	$\Delta\Delta G^{\text{dir}}$	$\Delta\Delta G^{\text{hyd}}$
Charge and radius			
CHARMM <sup>b</sup>	15.0 (1.0)	-4.6 (0.2)	19.6 (1.0)
PARSE <sup>c</sup>	17.3 (1.5)	-4.9 (0.2)	22.2 (1.6)
OPLS <sub>H</sub> <sup>d</sup>	12.5 (3.0)	-6.0 (0.1)	18.5 (3.0)
Radius only <sup>e</sup>			
ACCESS <sup>f</sup>	15.9 (1.0)	-4.6 (0.1)	20.5 (1.0)
ACCESS <sub>H</sub>	14.4 (0.7)	-4.6 (0.1)	19.0 (0.7)
CHARMM <sub>0</sub>	15.4 (0.9)	-4.6 (0.2)	20.1 (0.9)
CHARMM <sub>H</sub>	12.4 (0.3)	-4.9 (0.1)	17.3 (0.4)
OPLS	14.4 (0.3)	-4.8 (0.1)	19.1 (0.4)
OPLS <sub>H</sub>	12.3 (0.7)	-4.9 (0.1)	17.2 (0.7)
PARSE	16.6 (1.0)	-4.3 (0.2)	20.9 (1.0)

<sup>a</sup> All free energy values are in kcal/mol. Numbers in parenthesis indicate calculational uncertainty.

<sup>b</sup> Brooks et al. (1983).

<sup>c</sup> Sitkoff et al. (1994).

<sup>d</sup> OPLS ( $2^{-5/6}\sigma$  was used for the atomic radii; Jorgensen & Tirado-Rives, 1988) with a hydrogen radius of 1.25 Å.

<sup>e</sup> Radii from these parameter sets were used with the charges from the CHARMM parameter set. A subscript H indicates that a 1.25 Å hydrogen radius was used while a subscript 0 indicates a 0 Å hydrogen radius. The ACCESS and OPLS parameters give hydrogen a 0 Å radius so the parameters marked ACCESS and OPLS have a 0 Å hydrogen radius.

<sup>f</sup> Lee and Richards (1971); Eisenberg and McLachlan (1986).

uncertainty due to parameters. Naively, one might expect greater dependence on radii than charges because changes in radii exert their effect largely in the unbound state. The results show an overall range for the binding free energy from 12.3 to 17.3 kcal/mol, with a broader range for intramonomer contributions (17.2 to 22.2) than for intermonomer contributions (-4.3 to -6.0 kcal/mol). The broader range for intramonomer contributions is likely due to uncertainties in the desolvation energy. In other work, we have found that the implementation of the OPLS parameters used here tend to give larger direct terms and smaller desolvation penalties compared to CHARMM and PARSE, though it is not clear which is more correct (Hendsch et al., 1998). If the OPLS<sub>H</sub> full parameter set (which is purposely contrived to underestimate desolvation and overestimate interaction) is ignored, the intermonomer contributions range from -4.3 to -4.9 kcal/mol, which is very narrow. Here we note that changes to the radii, with fixed charges, affect  $\Delta\Delta G^{\text{hyd}}$  substantially more than  $\Delta\Delta G^{\text{dir}}$ . While uncertainties in the parameters prevent a precise value to be determined for the overall  $\Delta\Delta G^{\text{elec}}$  and, presumably, any single component or group of components, all parameter sets used produced the same overall result, that electrostatics are substantially destabilizing for complex formation in this system due to unfavorable intramonomer contributions that are roughly four times the size of favorable intermolecular interactions. Moreover, the checks that we have performed on a number of individual interactions with different parameter sets indicate that the identification of substantial vs. insubstantial contributions and their approximate relative magnitudes is independent over this set of parameters (results not shown). Thus, the results described here for the CHARMM parameter set provide a qualitatively accurate view of the relative magnitudes of individual and groups of contributions. While the actual values certainly depend in detail on what parameters are used, the general interpretation of the strong and weak, favorable and unfavorable interactions appears to be robust across parameter sets.

Four different molecular surfaces were tested to determine the effect of the representation of the surface on these calculations (Table 7). One surface used was the approximation to the molecular surface calculated by DELPHI v3.0 (which has been replaced by an improved algorithm in later releases; Gilson et al., 1988; Bharadwaj et al., 1995). The other three surfaces are based on the intersection of the exact representation of the molecular surface (calculated with a local program) with the grid lines (Richards, 1977; Connolly, 1983). For the unsmoothed surfaces (either approximated or calculated analytically), the dielectric constant is

**Table 7.** Dependence of  $\Delta\Delta G^{\text{elec}}$  and major components on the representation of the dielectric boundary<sup>a</sup>

Representation	$\Delta\Delta G^{\text{elec}}$	$\Delta\Delta G^{\text{dir}}$	$\Delta\Delta G^{\text{hyd}}$
DELPHI v3.0 surface <sup>b</sup>	19.4 (1.3)	-4.5 (0.1)	24.0 (1.3)
Exact surface <sup>c</sup>	15.0 (1.0)	-4.6 (0.2)	19.6 (1.0)
Smoothed <sup>d</sup>	13.6 (0.4)	-4.9 (0.1)	18.5 (0.4)
Modified smoothed <sup>e</sup>	13.6 (0.4)	-4.9 (0.1)	18.5 (0.4)

<sup>a</sup> All free energy values are in kcal/mol. Numbers in parenthesis indicate calculational uncertainty.

<sup>b</sup> Surface calculated by DELPHI v3.0.

<sup>c</sup> Exact representation of the molecular surface.

<sup>d</sup> Smoothing algorithm applied to exact surface.

<sup>e</sup> Modified smoothing algorithm applied to exact surface.

assigned to a grid line based on whether the center of the line is inside or outside the molecular surface. For the smoothed and modified smoothed surfaces, if a grid line crossed the molecular surface, the grid line was assigned a dielectric value between that of the interior and exterior dielectric depending on how much of the grid line was interior or exterior. The unmodified and modified smoothed surfaces differed in how they treated grid lines that crossed the surface more than once (Davis & McCammon, 1991; Mohan et al., 1992). The DELPHI v3.0 surface predicts the complex to be much more destabilizing than the surfaces based on the exact representation. Since the DELPHI v3.0 surface is an approximation of the exact surface used in the other calculations, it is assumed that this larger destabilizing effect is an artifact of this approximation and that the other surfaces better represent the true electrostatics of the system. The smoothed surfaces compute complex formation to be less destabilizing than the exact surface representation as expected from the theory (Mohan et al., 1992), but this difference is relatively small and unlikely to affect the interpretation of the results. The exact surface was used in the main results presented here.

## Discussion

The electrostatic contribution to docking for the GCN4 leucine zipper was computed to be unfavorable by 15 kcal/mol using a continuum electrostatic approach. That electrostatics disfavor binding in this system appears to be a robust result found for all parameter sets and other conditions for the calculation. This result for a peptide-peptide complex is consistent with continuum results for an antibody-antigen, a protein-DNA, a ligand-DNA, and an enzyme-inhibitor complex that also found electrostatics to be destabilizing (Misra et al., 1994a, 1994b; Sharp, 1996; Shen & Wendoloski, 1996).

The electrostatic decomposition used here to analyze GCN4 leucine zipper dimer formation provides a novel view of protein energetics. Whereas most discussions of molecular binding include both desolvation effects and new interactions formed in the complex, the continuum electrostatic treatment reveals that the overall hydration energy (defined as the change in the interactions of the collection of charges in each helix with their collective reaction field on binding) is not simply equal to the sum of the change in interaction of each functional group with its own reaction field on binding. Rather, there are a collection of cross terms involving changes to the interaction of each functional group with the reaction field due to other functional groups in the same helix; we have termed this an indirect interaction here, and it corresponds to solvent screened electrostatic interactions between groups. Clearly the assignment of an overall solvation energy to groupwise solvation and indirect components is a function of the groups chosen. Here we have used a functional-group based description; other descriptions, such as a residue-by-residue or even secondary-structure element based description (for larger applications), might be useful as well. However, an atom-by-atom description may be less useful, since interactions made by one atom in a functional group are frequently partially canceled by other atoms in the group (such as the two atoms in a carbonyl). The use of neutral groups, where appropriate, is probably preferable. While the hydration energy ( $= \Delta\Delta G^{\text{solv}} + \Delta\Delta G^{\text{indir}}$ ) was equal to 19.6 kcal/mol in the current study, the functional group decomposition used here gives a total  $\Delta\Delta G^{\text{solv}}$  of 24.0 kcal/mol and  $\Delta\Delta G^{\text{indir}}$  of -4.4 kcal/mol.

Thus, the difference between the overall desolvation penalty for the two helices and the sum of the desolvation penalties for each functional group was 22.4%. Although the desolvation penalty is generally unfavorable, in this study the indirect terms were net favorable, due largely to attractive intrahelical hydrogen-bonding interactions whose effective dielectric constant was decreased on binding. In other cases, such as the docking of proteins to DNA, the indirect term can be unfavorable due to phosphate-phosphate repulsions whose effective dielectric constant decreases upon binding.

An important finding is that intramolecular electrostatic interactions, particularly those near the binding interface, can be enhanced on binding due to reduced solvent screening. This is the  $\Delta\Delta G^{\text{indir}}$  contribution and includes 3 kcal/mol enhanced intrahelical interaction from the protein backbone (due largely to strengthening of the helical hydrogen bonds on binding). We have seen the importance of such interactions in a number of other complexes (Z.S. Hendsch, L.T. Chong, J.A. Caravella, & B. Tidor, unpubl. results) as well as in model systems for which we have designed charge-optimized ligands (Chong et al., 1998; Kangas & Tidor, 1998). Here we discuss experimental observations suggesting the importance of enhanced intramolecular effects on binding. Pauling and Corey (1953) first suggested that  $\alpha$ -helices could be distorted into coiled coils by a systematic pattern of shorter and longer helical backbone hydrogen bonds. Goodman and Kim (1991) were able to observe this distortion (following a heptad repeat) in which hydrogen bond lengths were systematically shorter at the buried face and longer at the exposed face of coiled coil helices. The electrostatic analysis presented here suggests that surface exposed  $\alpha$ -helices should have stronger helical hydrogen bonds along the buried face relative to the exposed. Structurally this differential could be expressed as a shortening of hydrogen bonds along the buried face relative to the exposed. Indeed, statistical studies of protein structures have revealed a distinct curvature of surface helices to form a more convex surface that is entirely consistent with this explanation (Blundell et al., 1983; Chakrabarti et al., 1986).

Not only was the overall electrostatics of binding computed to be destabilizing in this system, but the net effect of each and every close pairwise interaction formed across the dimer interface (salt bridges and hydrogen bonds) was also computed to be destabilizing. This extends our previous result, that salt bridges generally disfavor protein folding electrostatically, even though less desolvation penalty is generally incurred on protein binding than folding. We note, however, that we and others have found instances in which individual salt bridges do appear to favor binding (Xu et al., 1997).

The electrostatic decomposition employed here allows each functional group to be assigned an additive contribution to the electrostatics of binding. Operationally this corresponds to counting fully the solvation contribution of a group and counting half its interaction with other groups; the "half" effectively divides interactions equally between partners, which is appropriate due to reciprocity. Results here show that no group is more than marginally stabilizing (-0.2 kcal/mol), most groups are essentially neutral (within 0.5 kcal/mol of zero), and a few groups are substantially destabilizing. Interestingly, most of these destabilizing groups have been implicated in determining specificity in a variety of experiments (i.e., asparagine at the a position; Harbury et al., 1993; Potekhin et al., 1994; Gonzalez et al., 1996a, 1996b, 1996c; and charges at the e and g positions; O'Shea et al., 1992; John et al., 1994).

Scanning mutagenesis has been used as a technique to estimate experimentally the contribution of individual side chains to protein folding and binding (Clackson & Wells, 1995). Individual amino acid mutations are made (often to alanine), and the change in the folding or binding free energy is taken as some measure of that residue's contribution. Analogous manipulations can be carried out in the context of the computations carried out here, and they reveal certain limitations of such interpretations of scanning experiments. We use the decomposition to compute the effect of mutating a residue to its hydrophobic isostere (removing all partial atomic charge without allowing conformational relaxation beyond what is implied by the internal dielectric of four), analogous to scanning mutagenesis, and compare that to the computed electrostatic contribution to binding. The difference between the two is that, upon mutation a residue loses its full interactions with all other residues, whereas the contribution includes only one-half the strength of these interactions because their energetic effect is divided evenly between partners. When these two measures are plotted against one another (Fig. 4), one sees that very few contributions are favorable, but a large number of the mutational values are, suggesting that the approach used in scanning mutagenesis may overestimate the favorable energetic effect of a large number of side chains. Here, for instance, scanning mutational analysis would correctly predict that the fully exposed residues are relatively unimportant (b, c, and f positions), but incorrectly predict that interface charges (e and g positions) and the buried polar group (Asn16 at a) are stabilizing. Simply stated, scanning mutagenesis, while useful, has the disadvantage that adding up the energetic contribution from each position results in double counting all of the side-chain-side-chain coupling terms; when examined for individual residues, the coupling terms are overcounted relative to desolvation and other effects.

#### *Comparison with solvent effects determined by solvent accessible surface area*

It has been proposed by a number of researchers that the solvation penalty incurred on protein folding and binding can be reasonably approximated with a surface-area based potential (Eisenberg & McLachlan, 1986; Ooi et al., 1987). Here we investigate the model of Eisenberg and McLachlan (1986) in the light of our results. The solvation contribution in that model is equal to the change in solvent accessible surface area (Lee & Richards, 1971) multiplied by an atom-dependent atomic solvation parameter and summed over all atoms. Atomic solvation parameters have been determined for protein atoms using experimentally determined free energy of transfer values for partitioning of amino acids between water and organic solvents such as octanol (in which the octanol environment models the interior of a protein; Eisenberg & McLachlan, 1986). This type of potential, even though approximate, could be particularly useful because it is relatively rapid to evaluate and analytic derivatives are available for the accessible surface area calculation (Richmond, 1984), making it an inexpensive solvation alternative for molecular dynamics simulations (Wesson & Eisenberg, 1992).

Here we compare our solvation results computed with continuum electrostatics with the analogous results computed with the Eisenberg and McLachlan (1986) potential. To facilitate the comparison, the electrostatic contribution from the Eisenberg-McLachlan model was calculated as the difference between the solvation contribution computed normally and a computation in which all atoms had the atomic solvation parameter of carbon (but their ordinary

radius). This corresponds to the reference state used for the continuum electrostatic calculation (i.e., the free energy for introducing partial atomic charges into a hydrophobic version of a protein).

The Eisenberg-McLachlan model computes solvation on an atom-by-atom basis, with no cross terms representing effects of the solvation on a group due to the chemical nature of its neighbors; effects from neighbors enter only to the extent that they bury the atom in question. Note, however, that the model was parameterized using transfer measurements for amino acids so might contain correlation information to the level of intact amino acids but is not expected to include correlations beyond that to larger units of structure. Here we carried out continuum calculations on a functional group basis and explicitly included cross terms corresponding to the interactions of each functional group with the reaction field of other functional groups in the same helix. One test of the basis for an atomic-solvation-parameter model is the relative magnitude of these cross terms; a second important criterion is whether these indirect effects always have roughly the same relative size or whether they can vary substantially in proteins. As noted above, the indirect terms account for 22.4% of the hydration energy here. Another difference between the models is that even groups whose accessible surface area does not change often pay a desolvation penalty, which is neglected in the Eisenberg-McLachlan model. Here this amounts to only 1.3 kcal/mol.

The electrostatic desolvation penalty computed using the Eisenberg-McLachlan model is 10.7 kcal/mol, roughly half the hydration (19.6 kcal/mol) or solvation (24.0 kcal/mol) energy using a protein dielectric of four (Table 8). Interestingly, when a dielectric of 10.3, corresponding to octanol, is used for the protein, the continuum results match somewhat better (7.5 kcal/mol for hydration and 9.1 kcal/mol for solvation). While these overall values are in moderate to good agreement, a comparison of individual side-chain solvation penalties reveals large relative discrepancies between the two models (Fig. 5). Apparently, details of the shape of the dielectric boundary and relative placement of polar and charged groups lead to electrostatic solvation effects that are handled only approximately by a surface-area based method. Nevertheless, when summed over a large number of functional groups, this simple model performs remarkably well.

**Table 8.** Eisenberg-McLachlan solvation model compared with DELPHI<sup>a</sup>

	Total <sup>b</sup>	Side chain <sup>c</sup>	Carbonyl <sup>c</sup>	Amino <sup>c</sup>
Eisenberg-McLachlan	10.7	10.6	0.1	0.0
Desolvation only <sup>d</sup>				
DELPHI $\epsilon = 4$	24.0	19.5	4.0	0.5
DELPHI $\epsilon = 10.3$	9.1	7.7	1.3	0.1
Desolvation and indirect <sup>e</sup>				
DELPHI $\epsilon = 4$	19.6	18.0	1.4	0.1
DELPHI $\epsilon = 10.3$	7.5	6.9	0.5	0.1

<sup>a</sup> All energy values are in kcal/mol.

<sup>b</sup> The calculated total for all groups.

<sup>c</sup> The sum calculated for only the type of group specified.

<sup>d</sup> The sum of all desolvation penalties calculated by DELPHI.

<sup>e</sup> The sum of all desolvation and indirect penalties calculated by DELPHI. For indirect interactions occurring between groups of different types, half of the indirect interaction is assigned to the total for each group.

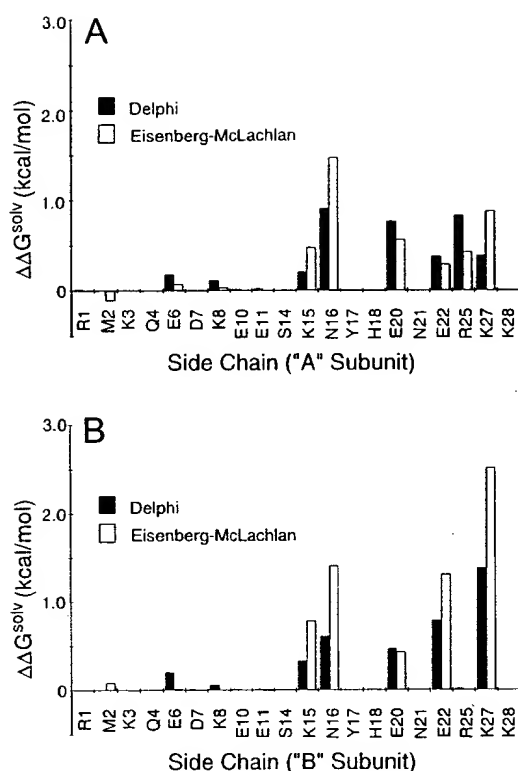


Fig. 5. A comparison of  $\Delta\Delta G^{\text{solv}}$  for all charged and polar side chains determined by a continuum electrostatic computation using an internal dielectric constant for the protein of 10.3 with the analogous calculation using the Eisenberg-McLachlan atomic solvation model (see text). (A) The A chain of the leucine zipper (as designated in PDB entry 2zta); (B) the B chain.

## Materials and methods

The electrostatic contribution to the free energy of complex formation ( $\Delta\Delta G^{\text{elec}}$ ) is defined here as the difference in total electrostatic energy between the bound and the unbound state:

$$\Delta\Delta G^{\text{elec}} = \Delta G_{\text{bound}}^{\text{elec}} - \Delta G_{\text{unbound}}^{\text{elec}} \quad (1)$$

The electrostatic energy of each state is represented as a  $\Delta G$  because it represents the difference in free energy between the actual complex and its hydrophobic isostere (a hypothetical complex with the same size and shape, but which is completely hydrophobic). For the bound state the coordinates of the crystal structure were used; for the unbound state, each helix was maintained in the same conformation as in the bound state but was treated as infinitely separated from the other helix. The total electrostatic free energy of each state was calculated as  $\sum_i \frac{1}{2} q_i \phi_i$  so the total electrostatic energy was

$$\Delta\Delta G^{\text{elec}} = \sum_{i \in C} \frac{1}{2} q_i \phi_i^{\text{C-bound}} - \left( \sum_{i \in A} \frac{1}{2} q_i \phi_i^{\text{A-unbound}} + \sum_{i \in B} \frac{1}{2} q_i \phi_i^{\text{B-unbound}} \right) \quad (2)$$

where  $C$  is the set of atoms in the complex,  $A$  is the set of atoms in helix A (one-half of the complex),  $B$  is the set of atoms in helix B (the other half of the complex),  $q_i$  is the partial atomic charge at atom center  $i$ , and the notation  $\phi_i^{y-z}$  is to be read as the potential at atom center  $i$  due to the charge(s)  $y$  in the  $z$  state.

Further calculations were then performed to allow this total electrostatic energy to be analyzed. When the linearized form of the Poisson-Boltzmann equation is used, superposition allows the total potential at any point to be expressed as a sum of contributions from individual charges or groups of charges. For an arbitrary group of atoms  $y$  in state  $z$ ,

$$\phi_i^{y-z} = \sum_{j \in y} \phi_i^{j-z} \quad (3)$$

Here this approach was used to give the contribution of individual functional groups. Each side chain, backbone carbonyl, and backbone  $\text{C}^\alpha\text{-N-H}$  was considered a separate group. At the N-terminus of each helix the acetyl cap was counted as one group, while at the C-terminus of each helix the terminal carboxylate was counted as one group. These were chosen because they represent chemically reasonable functional groups that each carry a net charge of 0, +1, or -1 with the charge parameters used. It should be noted that details of the results depend on this choice of groups. For example, the division between solvation and indirect terms is a function of how atoms are assigned to groups. The choice made here is based on one intuitive view of protein structure.

The contribution of an arbitrary group  $j$  to the total binding free energy was then calculated by summing one-half the charge at all atom positions times the potential at that atom due to group  $j$ ,

$$\Delta\Delta G_j^{\text{contrib}} = \sum_{i \in C} \frac{1}{2} q_i \phi_i^{j\text{-bound}} - \left( \sum_{i \in A} \frac{1}{2} q_i \phi_i^{j\text{-unbound}} + \sum_{i \in B} \frac{1}{2} q_i \phi_i^{j\text{-unbound}} \right) \quad (4)$$

Note that if group  $j$  is in helix A, the potential due to group  $j$  at the atoms of helix B will be zero, because the preformed helices are infinitely separated from one another in the unbound state. Summing the  $\Delta\Delta G_j^{\text{contrib}}$  terms for each group yields  $\Delta\Delta G^{\text{elec}}$ , the total electrostatic free energy of complex formation. Due to reciprocity, related expressions can be used to achieve the same quantity.

The contribution of each term was subdivided further into solvation, direct, and indirect terms. For a group  $j$  that is part of helix A, Equation 4 can be rearranged to yield

$$\Delta\Delta G_j^{\text{contrib}} = \sum_{i \in j} \left( \frac{1}{2} q_i \phi_i^{j\text{-bound}} - \frac{1}{2} q_i \phi_i^{j\text{-unbound}} \right) + \sum_{i \in B} \frac{1}{2} q_i \phi_i^{j\text{-bound}} + \sum_{i \in A; i \notin j} \left( \frac{1}{2} q_i \phi_i^{j\text{-bound}} - \frac{1}{2} q_i \phi_i^{j\text{-unbound}} \right) \quad (5)$$

The first term of this equation, in which the summation is only over the atoms of group  $j$ , represents the loss of solvent interactions of group  $j$  upon binding ( $\Delta\Delta G_j^{\text{solv}}$ ). The second term of the equation, which is a summation over the atoms of the other helix, is called the direct or intermolecular term ( $\Delta\Delta G_j^{\text{dir}}$ ) and is the



solvent screened coulombic interaction between group  $j$  and the charges in the other helix. The final term, which is summed over the charges in the same helix as group  $j$  except for the charges of  $j$ , is called the indirect or intramolecular term ( $\Delta\Delta G_j^{\text{indir}}$ ) and is the change in solvent screening between the bound and unbound state of group  $j$  and the other charges of that helix.

The direct and indirect terms were further divided into the direct and indirect interactions between a pair of groups rather than a single group with an entire helix by summing over only the atoms in the second group instead of the whole helix. For example, the direct interaction between two groups,  $j$  and  $k$ , is

$$\Delta\Delta G_{jk}^{\text{dir}} = \sum_{i \in k} \frac{1}{2} q_i \phi_i^{j\text{-bound}} + \sum_{i \in j} \frac{1}{2} q_i \phi_i^{k\text{-bound}} = \sum_{i \in k} q_i \phi_i^{j\text{-bound}}. \quad (6)$$

The interaction can be calculated either by summing the contribution of  $k$  to  $j$ 's total direct contribution and  $j$  to  $k$ 's total direct contribution or by dropping the factor of  $\frac{1}{2}$  from one or the other since the two terms are equivalent by reciprocity. From this description, it is clear that the interaction between two groups is divided evenly into the  $\Delta\Delta G^{\text{contrib}}$  of each group. Thus, the solvation term is due to changes in the interaction of a group with its own reaction field on binding, the direct term includes coulombic and reaction field interactions with charges in the other helix, and the indirect term reflects changes in interactions with the reaction field due to charges in the same helix.

The effect of mutating a single group to its hydrophobic isostere is not captured by the  $\Delta\Delta G^{\text{contrib}}$  because when a single group loses all partial atomic charge, it loses all of its interactions, rather than half. A separate term ( $\Delta\Delta G^{\text{mut}}$ ) is calculated as

$$\Delta\Delta G_j^{\text{mut}} = \Delta\Delta G_j^{\text{solv}} + 2\Delta\Delta G_j^{\text{dir}} + 2\Delta\Delta G_j^{\text{indir}}. \quad (7)$$

Note that it is inappropriate to sum  $\Delta\Delta G^{\text{mut}}$  terms for individual groups together because this leads to double counting of the interactions between groups.

For all calculations the coordinates of the crystal structure (Protein Data Bank (PDB) entry 2zta; O'Shea et al., 1991) were used from the PDB (Bernstein et al., 1977; Abola et al., 1987, 1996). Only the coordinates of the first 31 residues of the 33 residue peptide that was crystallized were located in this structure. Rather than attempt to build in the missing residues by modeling, only the first 31 residues were used for these calculations. The C-terminus was treated as charged, so that these calculations could be interpreted as the rigid docking of a 31 residue peptide with the same conformation as the 33 residue peptide. The effect of the charged C-terminus on the electrostatics of the complex was minimal (mutating them both to their hydrophobic isosteres would stabilize the complex by  $-0.03$  kcal/mol).

Polar hydrogens were built onto the crystal structure using the HBUILD facility (Brünger & Karplus, 1988) in CHARMM (Brooks et al., 1983). Arg, Lys, Glu, and Asp side chains were modeled in the charged form and His in neutral form. For the unbound state calculations each individual helix was placed on the grid in exactly the same way as for the bound state so that the energy resulting from the placement of the charges on the grid would be expected to cancel. Unless otherwise stated, all calculations were done using the linearized form of the Poisson-Boltzmann equation using a locally modified version of the continuum electrostatics program DELPHI (Gilson & Honig, 1987; Gilson et al., 1988; Sharp &

Honig, 1990) that was extended to use an exact representation of the molecular surface with a probe radius of  $1.4$  Å rather than the approximation used in this version (C.V. Sindelar and B. Tidor, unpubl. results). Tests were run with the nonlinear Poisson-Boltzmann equation and the total energy was found to differ by  $<0.1$  kcal/mol from the linearized form. All charges and radii were taken from the CHARMM PARAM19 (Brooks et al., 1983) polar-hydrogen parameter set. A dielectric constant of 80 was used for solvent and 4 was used for the protein. A salt concentration of  $0.145$  M was used with a  $2.0$  Å Stern layer (Bockris & Reddy, 1973; Gilson & Honig, 1987). Each calculation was done using "focusing" in which a low grid spacing calculation (using 23% fill and Debye-Hückel boundary conditions; Klapper et al., 1986) was done to determine the potential at the grid boundary for a higher grid spacing calculation (using 92% fill; unless stated otherwise, this results in a final grid spacing of  $1.06$  grids units per Å for the  $65 \times 65 \times 65$  grid used for most of the calculations). Each number reported is the average of 10 translations of the molecule relative to the grid, and the uncertainty is reported as twice the standard deviation of the mean (Taylor, 1982).

## Acknowledgments

We thank Peter S. Kim, Tom Alber, and members of the Tidor research group for helpful discussions and insightful comments. This work was supported by the National Institutes of Health (GM47678 and GM55758).

## References

- Abola EE, Bernstein FC, Bryant SH, Koetzle TF, Weng J. 1987. Protein Data Bank. In: Allen FH, Bergerhoff G, Sievers R, eds. *Crystallographic databases—Information content, software systems, scientific applications*. Bonn/Cambridge/Chester: Data Commission of the International Union of Crystallography. pp 107–132.
- Abola EE, Manning NO, Prilusky J, Stampf DR, Sussman JL. 1996. The Protein Data Bank: Current status and future challenges. *J Res Natl Inst Stand Technol* 101:231–241.
- Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol* 112:535–542.
- Bharadwaj R, Windemuth A, Sridharan S, Honig B, Nicholls A. 1995. The fast multipole boundary element method for molecular electrostatics: An optimal approach for large systems. *J Comput Chem* 16:898–913.
- Blundell T, Barlow D, Borkakoti N, Thornton J. 1983. Solvent-induced distortions and the curvature of  $\alpha$ -helices. *Nature* 306:281–283.
- Bockris JO'M, Reddy AKN. 1973. *Modern electrochemistry*. New York: Plenum.
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187–217.
- Brünger AT, Karplus M. 1988. Polar hydrogen positions in proteins: Empirical energy placement and neutron diffraction comparison. *Proteins Struct Funct Genet* 4:148–156.
- Chakrabarti P, Bernard M, Rees DC. 1986. Peptide-bond distortions and the curvature of  $\alpha$ -helices. *Biopolymers* 25:1087–1093.
- Chong LT, Dempster SE, Hendsch ZS, Lee L-P, Tidor B. 1998. Computation of electrostatic complements to proteins: A case of charge stabilized binding. *Protein Sci* 7:206–210.
- Clackson T, Wells JA. 1995. A hot spot of binding energy in a hormone-receptor interface. *Science* 267:383–386.
- Connolly ML. 1983. Analytical molecular surface calculation. *J Appl Cryst* 16:548–558.
- Crick FHC. 1953. The Fourier transform of a coiled-coil. *Acta Cryst* 6:685–689.
- Davis ME, McCammon JA. 1991. Dielectric boundary smoothing in finite difference solutions of the Poisson equation: An approach to improve accuracy and convergence. *J Comput Chem* 7:909–912.
- Eisenberg D, McLachlan AD. 1986. Solvation energy in protein folding and binding. *Nature* 319:199–203.



- Elcock AH, McCammon JA. 1996. The low dielectric interior of proteins is sufficient to cause major structural changes in DNA on association. *J Am Chem Soc* 118:3787–3788.
- Ellenberger TE, Brandl CJ, Struhl K, Harrison SC. 1992. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted  $\alpha$  helices: Crystal structure of the protein–DNA complex. *Cell* 71:1223–1237.
- Gilson MK, Honig BH. 1987. Calculation of electrostatic potentials in an enzyme active site. *Nature* 330:84–86.
- Gilson MK, Sharp KA, Honig BH. 1988. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J Comput Chem* 9:327–335.
- Gonzalez L Jr, Brown RA, Richardson D, Alber T. 1996a. Crystal structures of a single coiled-coil peptide in two oligomeric states reveal the basis for structural polymorphism. *Nature Struct Biol* 3:1002–1010.
- Gonzalez L Jr, Plecs JJ, Alber T. 1996b. An engineered allosteric switch in leucine-zipper oligomerization. *Nature Struct Biol* 3:510–515.
- Gonzalez L Jr, Woolfson DN, Alber T. 1996c. Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nature Struct Biol* 3:1011–1018.
- Goodman EM, Kim PS. 1991. Periodicity of amide proton exchange rates in a coiled-coil leucine zipper peptide. *Biochemistry* 30:11615–11620.
- Harbury PB, Zhang T, Kim PS, Alber T. 1993. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* 262:1401–1407.
- Hendsch ZS, Sindelar CV, Tidor B. 1998. Parameter dependence in continuum electrostatic calculations: A study using protein salt bridges. *J Phys Chem B* 102:4404–4410.
- Hendsch ZS, Tidor B. 1994. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci* 3:211–226.
- John M, Briand J-P, Granger-Schnarr M, Schnarr M. 1994. Two pairs of oppositely charged amino acids from Jun and Fos confer heterodimerization to GCN4 leucine zipper. *J Biol Chem* 269:16247–16253.
- Jorgensen WL, Tirado-Rives J. 1988. The OPLS potential function for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc* 110:1657–1666.
- Kangas E, Tidor B. 1998. Optimizing electrostatic affinity in ligand-receptor binding: Theory, computation, and ligand properties. *J Chem Phys* 109:7522–7545.
- Klapper I, Hagstrom R, Fine R, Sharp K, Honig B. 1986. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins Struct Funct Genet* 1:47–59.
- Landschulz WH, Johnson PF, McKnight SL. 1988. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science* 240:1759–1764.
- Lee B, Richards FM. 1971. The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol* 55:379–400.
- Lumb KJ, Kim PS. 1995. A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry* 34:8642–8648.
- Marten B, Kim K, Cortis C, Friesner RA, Murphy RB, Ringnalda MN, Sitkoff D, Honig B. 1996. New model for calculation of solvation free energies: Correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bonding effects. *J Phys Chem* 100:11775–11788.
- Misra VK, Hecht JL, Sharp KA, Friedman RA, Honig B. 1994a. Salt effects on protein–DNA interactions: The  $\lambda$ cl repressor and EcoRI endonuclease. *J Mol Biol* 238:264–280.
- Misra VK, Sharp KA, Friedman RA, Honig B. 1994b. Salt effects on ligand–DNA binding: Minor groove binding antibiotics. *J Mol Biol* 238:245–263.
- Mohan V, Davis ME, McCammon JA, Pettitt BM. 1992. Continuum model calculations of solvation free energies: Accurate evaluation of electrostatic contributions. *J Phys Chem* 96:6428–6431.
- Nautiyal S, Woolfson DN, King DS, Alber T. 1995. A designed heterotrimeric coiled coil. *Biochemistry* 34:11645–11651.
- Nicholls A, Sharp KA, Honig B. 1991. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct Funct Genet* 11:281–296.
- Oberoi H, Tripathi J, Yuan X, Allewell NM. 1996. Identification and analysis of long-range electrostatic effects in proteins by computer modeling: Aspartate transcarbamylase. *Proteins Struct Funct Genet* 25:300–314.
- Ooi T, Oobatake M, Némethy G, Scheraga HA. 1987. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc Natl Acad Sci USA* 84:3086–3090.
- O'Shea EK, Klemm JD, Kim PS, Alber T. 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* 254:539–544.
- O'Shea EK, Rutkowski R, Kim PS. 1992. Mechanism of specificity in the Fos–Jun oncoprotein heterodimer. *Cell* 68:699–708.
- Pauling L, Corey RB. 1953. Compound helical configurations of polypeptide chains: Structure of proteins of the  $\alpha$ -keratin type. *Nature* 171:59–61.
- Potekhin SA, Medvedkin VN, Kashparov IA, Venyaminov SY. 1994. Synthesis and properties of the peptide corresponding to the mutant form of the leucine zipper of the transcriptional activator GCN4 from yeast. *Protein Eng* 7:1097–1101.
- Prévost M. 1996. Concurrent interactions contribute to the raised  $pK_a$  of His 18 in barnase. *J Mol Biol* 260:99–110.
- Richards FM. 1977. Areas, volumes, packing, and protein structure. *Annu Rev Biophys Bioeng* 6:151–176.
- Richmond TJ. 1984. Solvent accessible surface area and excluded volume in proteins: Analytic equations for overlapping spheres and implications for the hydrophobic effect. *J Mol Biol* 178:63–89.
- Rick SW, Berne BJ. 1994. The aqueous solvation of water: A comparison of continuum methods with molecular dynamics. *J Am Chem Soc* 116:3949–3954.
- Sharp KA. 1996. Electrostatic interactions in hirudin–thrombin binding. *Biophys Chem* 61:37–49.
- Sharp KA, Honig B. 1990. Electrostatic interactions in macromolecules: Theory and applications. *Annu Rev Biophys Biophys Chem* 19:301–332.
- Shen J, Wendoloski J. 1996. Electrostatic binding energy calculation using the finite difference solution to the linearized Poisson–Boltzmann equation: Assessment of its accuracy. *J Comput Chem* 17:350–357.
- Sindelar CV, Hendsch ZS, Tidor B. 1998. Effects of salt bridges on protein structure and design. *Protein Sci* 7:1898–1914.
- Sitkoff D, Sharp KA, Honig B. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J Phys Chem* 98:1978–1988.
- Taylor JR. 1982. *An introduction to error analysis*. Mill Valley, California: University Science Books.
- Tidor B. 1994. Helix-capping interaction in a Cro protein: A free energy simulation analysis. *Proteins Struct Funct Genet* 19:310–323.
- Waldburger CD, Schildbach JF, Sauer RT. 1995. Are buried salt bridges important for protein stability and conformational specificity? *Nature Struct Biol* 2:122–128.
- Wang L, O'Connell T, Tropsha A, Hermans J. 1996. Energetic decomposition of the  $\alpha$ -helix-coil equilibrium of a dynamic model system. *Biopolymers* 39:479–489.
- Wesson L, Eisenberg D. 1992. Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci* 1:227–235.
- Wimley WC, Gawrisch K, Creamer TP, White SH. 1996. Direct measurement of salt-bridge solvation energies using a peptide model system: Implications for protein stability. *Proc Natl Acad Sci USA* 93:2985–2990.
- Xu D, Lin SL, Nussinov R. 1997. Protein binding versus protein folding: The role of hydrophilic bridges in protein associations. *J Mol Biol* 265:68–84.
- Yang A-S, Hitz B, Honig B. 1996. Free energy determinants of secondary structure formation: III.  $\beta$ -turns and their role in protein folding. *J Mol Biol* 259:873–882.
- Yang A-S, Honig B. 1995a. Free energy determinants of secondary structure formation: I.  $\alpha$ -helices. *J Mol Biol* 252:351–365.
- Yang A-S, Honig B. 1995b. Free energy determinants of secondary structure formation: II. Antiparallel  $\beta$ -sheets. *J Mol Biol* 252:366–376.
- Zeng X, Herndon AM, Hu JC. 1997. Buried asparagines determine the dimerization specificities of leucine zipper mutants. *Proc Natl Acad Sci USA* 94:3673–3678.

# Automated design of specificity in molecular recognition

James J. Havranek<sup>1</sup> and Pehr B. Harbury<sup>2</sup>

Published online 2 December 2002; doi:10.1038/nsb877

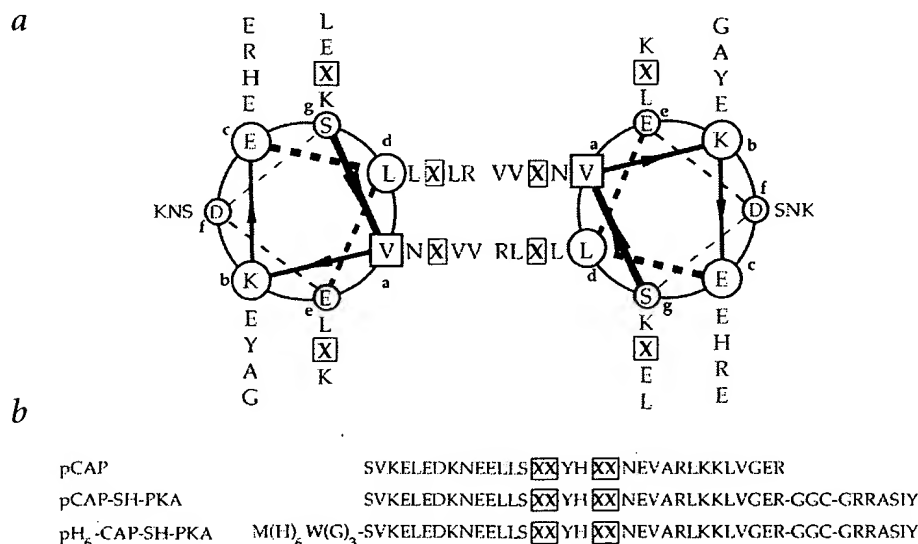
**Specific protein–protein interactions are crucial in signaling networks and for the assembly of multi-protein complexes, and represent a challenging goal for protein design. Optimizing interaction specificity requires both positive design, the stabilization of a desired interaction, and negative design, the destabilization of undesired interactions. Currently, no automated protein-design algorithms use explicit negative design to guide a sequence search. We describe a multi-state framework for engineering specificity that selects sequences maximizing the transfer free energy of a protein from a target conformation to a set of undesired competitor conformations. To test the multi-state framework, we engineered coiled-coil interfaces that direct the formation of either homodimers or heterodimers. The algorithm identified three specificity motifs that have not been observed in naturally occurring coiled coils. In all cases, experimental results confirm the predicted specificities.**

Computational protein design provides a rigorous test of our understanding of proteins. In effect, a design algorithm translates our hypotheses about protein structure and function into amino acid sequences. Experimental analysis of these sequences reports on the validity of the hypotheses. Recent design efforts have resulted in the realization of a novel backbone fold<sup>1</sup>, the redesign of a folding pathway<sup>2</sup> and the design of a zinc finger domain that does not require metal binding for stability<sup>3</sup>.

Most design studies follow the ‘inverted-folding’ strategy in which an optimal sequence for a preexisting backbone is selected by the design algorithm<sup>4</sup>. Protein structure is represented by a fixed backbone and a rotamer-based description of side chain conformation<sup>5</sup>. Amino acid sequences are selected that mini-

mize a potential energy function when computationally modeled in the target conformation. We refer to this procedure as ‘single-state’ design. The potential energy functions used in protein design include empirically weighted contributions derived from molecular mechanics potentials, secondary structure propensities, structural database statistics and surface-area scaled terms that depend on hydrophobic/polar (H/P) character<sup>6,7</sup>. Because they combine a diverse set of energetic and statistical considerations, we refer to these as ‘hybrid’ potential energy functions. This general approach has led to numerous impressive results from several groups<sup>2,3,8–13</sup>.

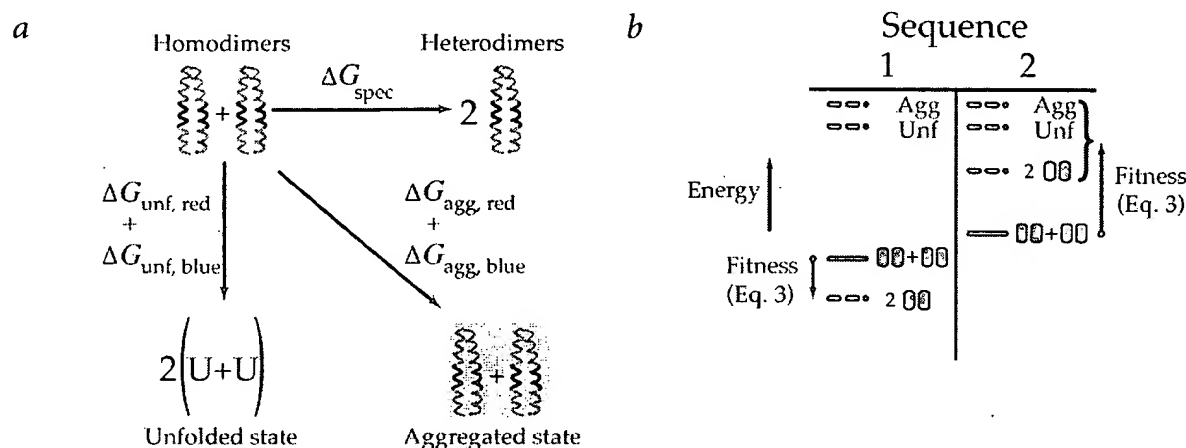
These successes suggest that the automated stabilization of fixed structures may be considered a solved problem. However,



**Fig. 1** A GCN4-derived scaffold for coiled-coil design. **a**, A helical wheel diagram of the pCAP sequence. Positions allowed to vary in the design calculation are denoted by ‘X’. These positions form the interface of the central heptad. The pCAP sequence is identical to an N-terminally capped variant of GCN4 (ref. 60) with the asparagine at position 16 shifted by one heptad level to position 9. **b**, Constructs used for the experimental characterization of designed sequences. ‘H<sub>6</sub>’ denotes a (His)<sub>6</sub>-tag; ‘SH’, a (Gly-Gly-Cys) linker; and ‘PKA’, a protein kinase A-tag.

<sup>1</sup>Biophysics Program and <sup>2</sup>Department of Biochemistry, Stanford University, Stanford, California 94305, USA.

Correspondence should be addressed to P.B.H. e-mail: harbury@cmgm.stanford.edu



**Fig. 2** Positive and negative design states. **a**, Schematic representation of competing states included in the design calculations targeting the homodimer conformation. Competing states are included to enforce homospecificity (upper right), solubility (lower right) and stability (lower left). The aggregated state is modeled as the target conformation embedded in a medium with a dielectric constant of 65. Each sequence considered by the genetic algorithm is subjected to conformational optimization in each of the four states. The fitness score for a given sequence is the transfer free energy from the target state (homodimer) to the ensemble of competing states (heterodimers, aggregated state and unfolded state).  $\Delta G_{\text{spec}}$  and  $\Delta G_{\text{unf}}$  are transfer free energies between states and can be measured experimentally. **b**, Energy diagram for two protein sequences in different conformational or associative states. A solid green line indicates the energy of the target state, the coiled-coil homodimers. The dashed magenta lines indicate the energies of competing states, including the unfolded protein, aggregated protein and the coiled-coil heterodimers. Sequence 1 minimizes the energy of the target state and would be incorrectly selected by the single-state design algorithm. The single-state design algorithm cannot recognize that the heterodimers are more stable than the homodimers because stabilities are computed only for the target state. The multi-state algorithm would correctly select sequence 2, because its transfer free energy from the target state to the ensemble of competing states is more positive than for sequence 1 (Eq. 3).

single-state approaches do not explicitly address discrimination between multiple states, which is a central feature of molecular specificity. Examples include proteins that selectively bind one small molecule without binding chemically related compounds, allosteric proteins that change conformation in the presence of a regulatory ligand and enzymes capable of binding transition states more tightly than ground states. To maximize specificity for a target state, the design algorithm must both stabilize the desired physical result (positive design) and destabilize undesired conformations, arrangements or states (negative design).

We present a general method for the automated design of specificity in molecular recognition. Following previous work<sup>1,14,15</sup>, we represent each design requirement as a separate state that the protein can adopt. The algorithm achieves specificity by selecting sequences calculated to have an energetic preference for the target state over the negative design states. We refer to this procedure as 'multi-state' design. Using a coiled-coil model system for molecular recognition, we show that the use of multiple states in our calculations is necessary. The multi-state algorithm discovers new specificity motifs unreported in naturally occurring coiled coils.

### Design of specific coiled-coil interfaces

We chose a dimeric coiled coil as our design scaffold because it represents the simplest protein-protein interface. Coiled coils have a characteristic heptad repeat (a–g, Fig. 1a). Positions a and d are typically occupied by hydrophobic residues, positions e and g by charged residues and positions b, c and f by polar residues. We redesigned positions a, d, e and g in the central heptad of the prototypical and well-studied homodimeric coiled coil GCN4 (ref. 16). Eight residues (four per helix) were varied, generating two distinct sequences. All non-proline amino acids were considered at the designed positions, allowing for a total of  $8 \times 10^9$  possible sequence outcomes.

A design intended to select two coiled-coil sequences that preferentially associate into homodimers and do not cross-hybridize with each other is illustrated (Fig. 2). Four states were modeled. The first state is defined as the folded homodimer conformation, which is the target state. The second state is the folded heterodimer conformation, which is included as a competing state to select against sequences that cross-hybridize. The third state is the unfolded state of the polypeptides, which is included as a competitor to select against sequences that are unstable. The fourth state is the aggregated state, which is included as a competitor to select against sequences with poor water solubility.

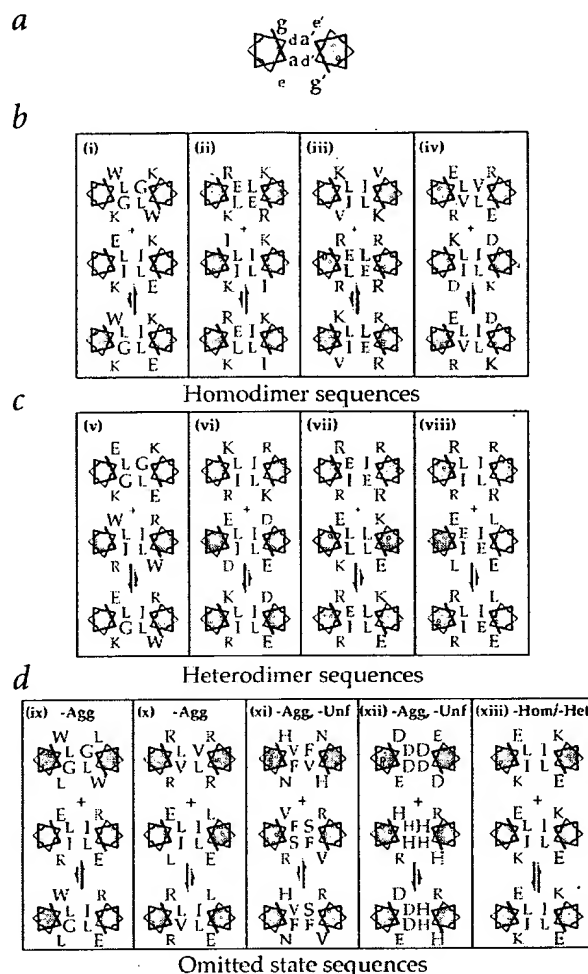
Free energies were evaluated for candidate sequences in each of the four states. The fitness of a sequence was defined as its computed transfer free energy from the target state to the ensemble of competing states (Fig. 2b). Single-state design algorithms select sequences with the lowest computed energy in the target state (sequence 1, Fig. 2b). The multi-state design algorithm selects sequences that maximize the fitness (sequence 2, Fig. 2b), ensuring specificity towards the target state.

A genetic algorithm was used to evolve a population of sequences that maximize the transfer free energy from the target state to the ensemble of competing states. To distinguish different classes of solutions within the population, we clustered the 100 sequences of highest fitness into four groups using BLASTClust<sup>17</sup>. The sequence with the largest transfer free energy from each cluster is reported (Table 1; Fig. 3).

### Identifying specific pair interactions

The designed sequences incorporate both previously identified and new amino acid motifs. To identify the pairwise interactions in these motifs that are responsible for the computed specificity, computational double-mutant cycles were performed. In the cases of sequences iv and vi (Fig. 3), specificity was achieved by patterning charged residues on the protein surface, which occurs naturally in coiled coils<sup>18</sup> and has been used in protein engineer-

**Fig. 3** Results of design calculations. **a**, Positions of designed residues in the central heptad of pCAP (Fig. 1). **b**, Designed homodimer sequences. The arrangement of residues for both the homodimer and heterodimer species is shown for each sequence pair. The target state of the design is indicated by the direction of the equilibrium arrow. Single letter abbreviations are used for the amino acids. Basic and acidic residues are shown in blue and red, respectively. **c**, Designed heterodimer sequences. The arrangement of residues for both the homodimer and heterodimer species is shown for each sequence pair. **d**, Omitted state sequences. Sequences result from calculations omitting one or more competing states (Table 1). The competing states omitted from each calculation are indicated at the top of the panel.



ing studies<sup>19</sup>. Several novel sequence patterns also emerge. Volume complementarity between a Trp side chain and a Gly side chain confers specificity in sequences i and v (Fig. 4a). Poor packing between a Leu side chain at heptad position a against  $\beta$ -branched side chains at positions g' and a' accounts for the homospecificity of sequence ii. In sequences iii and vii, a Glu side chain at heptad position d favors a basic amino acid at position e' over a hydrophobic alternative (Fig. 4b). Because position d of the heptad repeat is located in the hydrophobic core of the coiled coil, these sequences contain buried polar residues computationally engineered to confer specificity.

### Multiple design goals require multiple states

To test whether multiple states are required to achieve our design criteria, we performed a second set of calculations in which one or more competing states were omitted (sequences ix–xiii, Fig. 3d; Table 1). The calculations with limited sets of competitors demonstrate that the neglect of any state yields inferior results relative to the results obtained with the full set of competitors. The omission of the aggregated state gives rise to sequences with fewer charged residues (compare sequences ix and x with i and vi). Although it is not clear whether the aggregated state is required for the success of our design, the loss of polar residues at surface positions is generally undesirable. Designs lacking both the aggregated and unfolded states lead to sequence pairs predicted to be specific but also unstable (sequences xi and xii). When

**Table 1** Calculated (using the OPLS-UA potential) and observed thermodynamic quantities for designed sequence pairs

Sequence pair (A/B)	States <sup>1</sup>	$\Delta G_{\text{Fitness}}$ (kcal mol <sup>-1</sup> ) <sup>2</sup>	$\Delta G_{\text{spec,calc}}$ (kcal mol <sup>-1</sup> ) <sup>3</sup>	$\Delta G_{\text{spec,obs}}$ (kcal mol <sup>-1</sup> )	$\Delta G_{\text{unf,calc}}$ (kcal mol <sup>-1</sup> ) <sup>4</sup>	$\Delta G_{\text{unf,obs}}$ (kcal mol <sup>-1</sup> )	$\Delta G_{\text{agg,calc}}$ (kcal mol <sup>-1</sup> ) <sup>5</sup>
i WGLK/EILK	* C C C	+4.3	+9.7	+0.9	+1.8/+2.9	-5.4/+1.6	+4.8/+5.1
ii RLEK/IILK	* C C C	+3.8	+7.1	+3.0	+1.4/+2.1	-3.8/+1.9	+4.6/+4.6
iii KILV/RLER	* C C C	+3.8	+7.6	+1.6	+2.7/+1.6	-0.5/-3.6	+4.3/+4.5
iv EVLR/KILD	* C C C	+3.7	+4.8	+2.7	+3.1/+1.1	-0.9/-1.9	+4.9/+4.6
v EGLK/WILR	C * C C	+4.7	-12.1	-1.7	+3.0	-0.6	+4.8
vi KILR/EILD	C * C C	+4.7	-8.7	-2.5	+2.9	-0.9	+4.8
vii RIER/ELLK	C * C C	+4.5	-5.6	-2.5	+2.9	+1.7	+4.7
viii RILR/EIEL	C * C C	+4.4	-12.2	-1.7	+2.3	-1.6	+4.6
ix WGLL/EILR	* C C -	+5.7	+12.8		+3.4		+4.9/+3.9
x RVLR/EILL	C * C -	+6.4	-6.8		+3.9		+4.4
xi HFVN/VSFR	* C - -	+50.2	+50.2		-16.7		+4.6/+4.1
xii DDDE/HHHR	C * - -	+109.3	-109.3		-10.4		+4.9
xiii EILK/EILK	* - C C	+5.7	0.0		+2.9		+5.1/+5.1
EILK/EILK	- * C C	+5.7	0.0		+2.9		+5.1

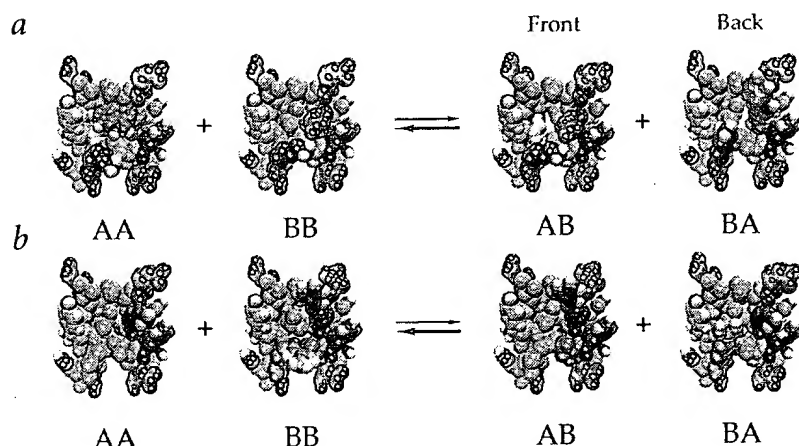
<sup>1</sup>The target state for each calculation is denoted with an asterisk, competing states with a 'C', and states omitted from the calculation with a minus sign. The abbreviations for the states are Hom, homodimers; Het, heterodimers; Unf, the unfolded state; and Agg, the aggregated state.

<sup>2</sup>Transfer free energy from target state to ensemble of competitors.

<sup>3</sup> $\Delta G_{\text{spec}}$  is defined as the free energy change when the two homodimers are rearranged to form the two heterodimers.

<sup>4</sup> $\Delta \Delta G_{\text{unf}}$  is defined as the free energy difference between the unfolded and target states, subtracted by the same value for the pCAP (KVLE / KVLE) sequence. For homodimer species,  $\Delta \Delta G_{\text{unf}}$  is reported for both sequences (A/B).

<sup>5</sup> $\Delta G_{\text{agg}}$  is defined as the free energy difference between the aggregated and target states. For homodimer species,  $\Delta G_{\text{agg}}$  is reported for both sequences (A/B).



**Fig. 4** Predicted specificity motifs. **a**, Sequence pair i (Table 1). In the AA homodimer, a Trp residue (green) occupies a space created by a Gly residue on the opposite helix (white). In the AB heterodimer, a Glu residue is opposite the Gly residue. The Glu residue extends into solvent, leaving a cavity in the hydrophobic core of the coiled coil. The Trp residue is placed opposite an Ile residue and cannot pack into the core. **b**, Sequence pair iii. In the BB homodimer, a Glu residue at the heptad **d** position (orange) is in close contact with an Arg residue at the opposing **e'** position (cyan). In the AB heterodimer, the Glu is opposite an uncharged residue (Val). Figures were generated with VMD<sup>61</sup>.

the homodimer competitor is omitted from the heterodimer design, all association specificity is lost (sequence xiii). Omission of the heterodimer competitor from the homodimer design results in the same loss of specificity (sequence xiii). We conclude that both positive and negative design states must be considered to achieve specificity in our calculations.

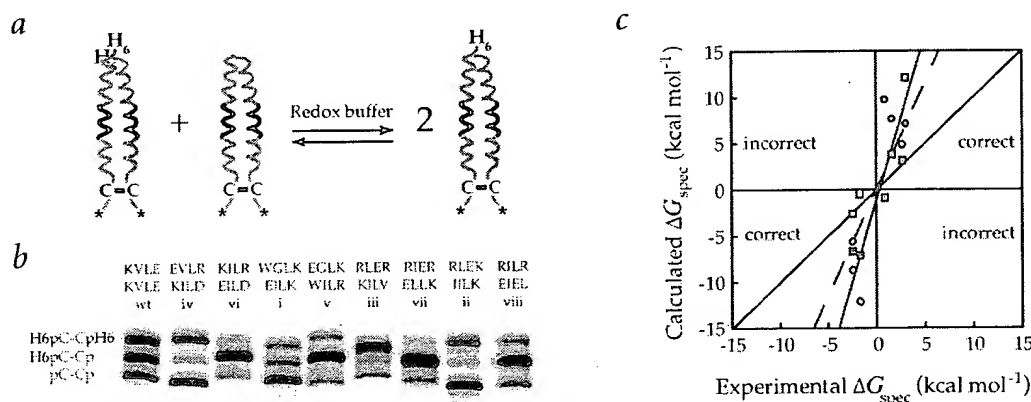
#### Experimental validation of designed sequences

To test the energetic predictions, sequence pairs i–viii were expressed, purified and characterized experimentally. We first determined whether the target species form parallel dimeric structures by measuring whether the apparent melting temperatures ( $T_m$ ) of C-terminally disulfide-bonded–target coiled coils vary with peptide concentration<sup>18</sup>. All of the  $T_m$ s were observed to be concentration independent. These data rule out the possibility that the disulfide-bonded coiled coils form higher order oligomers or adopt antiparallel conformations. For six of the species, the dimer oligomerization state was confirmed independently by analytical ultracentrifugation.

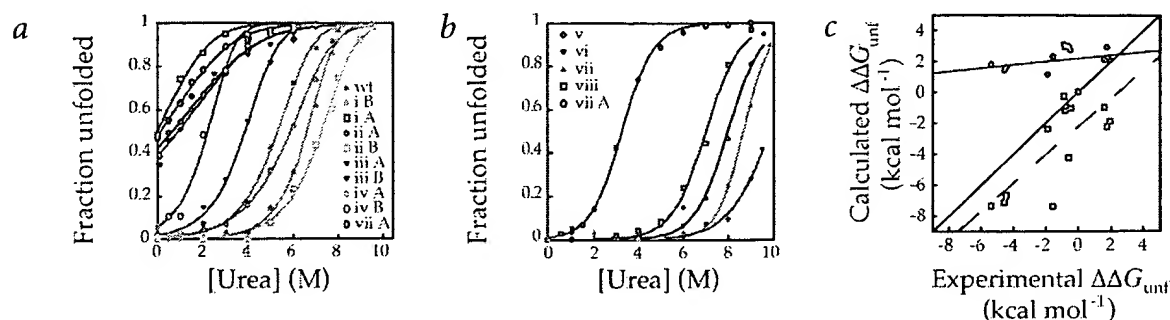
A disulfide-exchange assay was used to measure directly the equilibrium between the homodimer and heterodimer states of the designed coiled coils (Fig. 5a,b). In each instance, specificity

for the desired association state was achieved (Table 1; Fig. 5c). Two sets of predictions are shown. The first set is computed with the OPLS-UA potential energy function, which was used in the design calculation<sup>20</sup>. In addition, we report specificities calculated identically, using the CHARMM19 potential energy function<sup>21</sup>, which was applied after the design to evaluate the selected sequences. Both sets of predictions correlate with the measured values (the square of the correlation coefficient ( $R^2$ ) is 0.7 for both OPLS-UA and CHARMM19).

To test our predictions of unfolding free energies, stabilities were measured for sequences i–viii by urea denaturation. All measurements were taken in 5 mM phosphate buffer, consistent with the low salt environment used for the design calculation. For homodimer species, melts of unmodified coiled coils were performed. For heterodimer species, disulfide-bonded coiled coils were studied to prevent the formation of a mixed population of dimers. The data were fit assuming a two-state bimolecular (homodimers) or unimolecular (heterodimers) folding reaction (Fig. 6a,b). Stabilities were extracted from the data and referenced to that of the pCAP peptide, the parental sequence for the design calculation. We compared the stabilities predicted using the OPLS-UA<sup>20</sup> and CHARMM19 potential energy func-



**Fig. 5** Specificity of designed coiled coils. **a**, One member of each sequence pair was expressed in *E. coli* with an N-terminal His<sub>6</sub>-tag (red), whereas the other was expressed without the tag (blue). The proteins were allowed to exchange helix partners in the presence of redox reagents, which facilitated the breaking and reforming of disulfide bonds, until equilibrium was reached. The exchange reaction was then quenched, radioactively labeled and analyzed by SDS-PAGE. **b**, Autoradiograph of electrophoretically separated exchange reaction. The top band is composed of His<sub>6</sub>-tagged homodimers; the middle band, heterodimers; and the bottom band, untagged homodimers. Sequence pairs are colored corresponding to whether they were expressed with (red) or without (blue) a His<sub>6</sub>-tag. **c**, Specificities calculated using the OPLS-UA potential energy function<sup>20</sup> (blue circles; slope = 3.5 and  $R^2 = 0.7$ ) or the CHARMM19 potential energy function<sup>21</sup> (red squares; slope = 2.3 and  $R^2 = 0.7$ ) plotted against the measured values. Lines of best fit are shown in solid blue (OPLS-UA) and dashed red (CHARMM19). The diagonal is shown as a solid black line. The quadrants of the graph are labeled to indicate where the computation correctly predicts measured specificity.



**Fig. 6** Stability of designed coiled coils. **a**, Urea denaturation of designed homodimer coiled coils. Sequences are identified by lower case roman numerals (Table 1). The pCAP sequence is denoted 'wt'. Curves were fit using a two-state bimolecular model. **b**, Urea denaturation of disulfide-bonded dimers of designed heterodimer coiled coils. Curves were fit using a two-state unimolecular model. Sequences are identified by lower case roman numerals (Table 1). **c**, The calculated stability for each of the designed species (eight homodimers, four heterodimers and the scaffold) is plotted against the observed value. Blue circles denote values calculated using the OPLS-UA potential energy function<sup>20</sup>, and red squares are those using the CHARMM19 potential energy function<sup>21</sup>. Lines of best fit are shown in solid blue (OPLS-UA; slope = 0.1,  $R^2 = 0.1$ ) and dashed red (CHARMM19; slope = 0.9,  $R^2 = 0.6$ ). The diagonal is shown as a solid black line.

tions<sup>21</sup> with the experimental stabilities. All target species are predicted to be more stable than the pCAP peptide by the OPLS-UA potential energy function. In contrast, stabilities calculated with the CHARMM19 potential are in closer agreement with the observed values.

## Discussion

Our design algorithm differs from previous efforts<sup>2,3,8,9</sup> in two ways. First, we select sequences that maximize the transfer free energy of a protein from a target state to an ensemble of explicitly represented competitors, rather than optimizing the computed potential energy for a single target state. As a result, sequence optimization is necessarily distinct from structural optimization, and structural optimization must be performed separately for each state<sup>22</sup>. The computed specificities are used to guide the subsequent sequence search. Second, we evaluate conformational free energies using a standard molecular mechanics potential energy function with a continuum solvent model (MM/CS) rather than a hybrid energy function. This allows us to directly compare predicted and observed free energies. Although MM/CS potential energy functions are still under development, they are more thoroughly parameterized and tested than hybrid energy functions. Because they can account for the energetics of small molecules, nucleic acids and proteins, they are expected to be more general than hybrid energy functions. Finally, the use of a standard molecular mechanics potential energy function allows for its modular substitution with improved potentials, as advances (such as polarizable potential energy functions) emerge from other fields of computational chemistry.

## Specificity in protein design

Specificity is, by definition, a multi-state property<sup>23</sup>. The probability that a designed protein will adopt a target conformation or state is given by:

$$P_{\text{target}} = (e^{-A_{\text{target}}/RT}) / (\sum_{\text{target, competitors}} e^{-A/RT}) \quad (1)$$

where  $A_{\text{target}}$  is the free energy of the target conformation, and  $A$  is the free energy of each conformation in the sum in the denominator.

Single-state design is predicated on maximizing the numerator of Eq. 1, neglecting the effects of sequence variation on the denominator (the partition sum). This approach has been used

successfully to engineer specificity<sup>15,24</sup>. However, if any competitor conformation structurally resembles the target conformation, optimization for the target will be correlated with optimization for the competitor, and the single-state strategy will likely break down<sup>14,25,26</sup>.

Failure of single-state design is observed in lattice models of proteins, resulting in heteropolymer sequences that fold into multiple conformations<sup>27</sup>. To address this deficiency, a new generation of lattice-design algorithms selects sequences that directly optimize  $P_{\text{target}}$  in Eq. 1 rather than target stability<sup>28–30</sup>. Although optimizing  $P_{\text{target}}$  may seem computationally prohibitive, given the large number of states that could contribute to the denominator in Eq. 1, it has been noted<sup>31,32</sup> that the partition sum is dominated by a small number of low-energy conformations that are structurally similar to the target. The partition sum can thus be approximated by modeling this subset of near-native conformations.

The differences between the single-state and multi-state strategies are highlighted by the manner in which they achieve the 'hydrophobic in/polar out' pattern observed in naturally occurring proteins. This pattern reflects the realities that buried charges can be destabilizing and that an excess of hydrophobic residues at the surface can lead to aggregation. In single-state design algorithms, the unfolded and aggregated states are not modeled. Consequently, the selection of charged residues at buried positions is discouraged by penalizing the burial of polar surface area, by excluding polar residues from buried positions<sup>7,12</sup> or by constraining amino acid composition<sup>33</sup>. Likewise, hydrophobic residues are often excluded from consideration at surface positions to prevent aggregation. Although sequence constraints may be expedient for enforcing 'hydrophobic in/polar out' patterning, protein function often depends on exceptions to this rule<sup>34–40</sup>. In the multi-state approach, the patterning of hydrophobic and polar residues arises as a natural consequence of simultaneous competition against the unfolded and aggregated states. Thus, polar residues are not excluded from the cores of proteins; their selection is based on an energetic balance between the requirements for stability and specificity.

The multi-state approach only offers an advantage for designing against undesired competitor states that are known and can be modeled. With respect to unknown competitors, the single-state and multi-state approaches are equivalent. One must

hope that specificity against unknown competitors will arise fortuitously as a consequence of sequence optimization for the target state. To assess the magnitude of such fortuitous specificity, we measured  $\Delta G_{\text{spec}}$  for seven pairs of homodimer sequences that were not deliberately designed to disfavor cross-hybridization with each other (iA/iiiB, iA/ivB, iiA/iiiB, iiA/ivB, iiiA/ivB, ivA/iiB and ivA/iiiB; Fig. 3)). The  $\Delta G_{\text{spec}}$  values for these pairs range from  $-0.1$  to  $1.1 \text{ kcal mol}^{-1}$ , averaging  $0.6 \text{ kcal mol}^{-1}$  (data not shown). Seven of the eight engineered sequence pairs (sequences i–viii, Fig. 3; Table 1) show values of  $\Delta G_{\text{spec}}$  exceeding in magnitude the largest value of  $\Delta G_{\text{spec}}$  that arises fortuitously.

### Assessment of the physical model

Comparison of experimentally measured free energies with predicted values reports on the accuracy of the physical model used for design, which includes the side chain rotamer library, the backbone representation and the potential energy function. The results demonstrate that MM/CS energy potentials are capable of conferring functional specificity on designed proteins. However, the quantitative agreement between energetic predictions and measured values leaves room for improvement.

As observed by others, we find that the accuracy of energetic estimates strongly depends on the number of rotamers used to model side chain conformations<sup>41</sup>. The library we used included 1,064 rotamers to represent the 19 non-proline amino acids. Optimizing the side chain coordinates in the presence of the fixed backbone after rotamer placement was necessary to achieve energy values that could be compared across all states<sup>9,42</sup>.

Our physical model does not consider backbone movements of the protein, which could mitigate unfavorable interactions in the negative design states. This limitation probably contributes to the overestimation of the specificities by both the OPLS-UA<sup>20</sup> and CHARMM19 potential energy functions<sup>21</sup> (Fig. 5c). Incorporating backbone flexibility in the multi-state framework should be possible by including several fixed backbone conformations for each of the competing states.

All potential energy functions contain errors. The design process likely inflates the cumulative error in the force field used for the design calculation. Presumably, the genetic algorithm selects sequences for which errors in the OPLS-UA potential energy function are correlated and preferentially stabilize the target state. In contrast, the designed sequences are expected to sample errors that are present in the CHARMM19 potential energy function randomly, yielding a more accurate assessment of their stabilities. The differences in the OPLS-UA and CHARMM19 stability predictions derive from the bonded and Lennard-Jones terms in the potential energy function (data not shown). The results suggest that cross-validation by independent potential energy functions could be used to identify designed sequences that likely contain accumulated errors before time-consuming experimental efforts are initiated.

### Conclusion

We have presented a general method for incorporating specificity into protein design. Each positive and negative design requirement is embodied as a separate state in our algorithm. We have verified experimentally that this multi-state framework produces functionally specific protein–protein recognition. The use of a molecular mechanics potential energy function with a continuum solvent model allows for comparison of the predicted and observed free energies. The results suggest that the use of several potential energy functions may help to minimize the effects of errors present in these functions.

Our framework for multiple competing states is applicable beyond the simple protein–protein interactions that we have considered here. For example, larger sets of orthogonal coiled coils could be designed to direct complex self-assembly processes. Competing states in which a protein is bound to ‘decoy’ ligands could be used to direct the design of specific small-molecules (for example, see ref. 14). Finally, an explicit framework for the stabilization of the transition state of a reaction relative to its ground states should be possible<sup>43</sup>.

### Methods

**Rotamer library and optimization.** Backbone coordinates for a symmetric idealized coiled coil were generated from a mathematical model using parameters optimal for Val and Leu residues at heptad positions **a** and **d**<sup>44</sup> (Fig. 1). The most commonly occurring rotamers for an  $\alpha$ -helix were taken from the backbone-dependent library of Dunbrack and Karplus<sup>45</sup>. Sufficient rotamers for each amino acid were extracted to account for 95% of all observed conformations. Sulfhydryl and hydroxyl hydrogens were added with dihedral angles of  $-60^\circ$ ,  $60^\circ$  and  $180^\circ$ . Rotamers were built onto the backbone structure and energy minimized using either the OPLS-UA<sup>20</sup> or CHARMM19 (ref. 21) geometric and van der Waals potential energy terms and a  $20^\circ$  square-well dihedral restraint<sup>20</sup>. Additional rotamers were then introduced, offset from the minimized values by 1.3 s.d. in the  $\chi_1$  dihedral angle<sup>45</sup> for lysine, methionine, glutamine, glutamate and arginine, and in  $\chi_1$  and  $\chi_2$  for all other amino acids ( $20^\circ$  for the hydrogen dihedrals above). Atom positions for additional rotamers were energy minimized, with their dihedral angles held fixed. Rotamer probabilities were optimized following Koehl and Delarue<sup>46</sup>. Because the mean-field algorithm does not guarantee convergence to the global minimum<sup>47</sup>, each sequence in Table 1 was repacked 50x with different random initial rotamer probabilities. The results agreed to within  $0.01 \text{ kcal mol}^{-1}$ , suggesting that the repacking algorithm identifies the globally optimal conformation for these sequences.

**Energy function.** The potential energy of the system is approximated in a pairwise factorable form<sup>48</sup> and decomposed into the following contributions:

$$U_{\text{total}} = U_{\text{Geom}} + U_{\text{LJ}} + U_{\text{MTK}} / \gamma \quad (2)$$

$U_{\text{Geom}}$  consists of the bonded energy terms from the OPLS-UA<sup>20</sup> or CHARMM19 (ref. 21) force field.  $U_{\text{MTK}} / \gamma$  is identical to the FDPB /  $\gamma$  solvation energy<sup>48</sup>, with the electrostatic energies calculated from PARSE parameters<sup>48</sup> using the modified Tanford-Kirkwood algorithm<sup>49</sup>. The solvent and protein dielectric constants used were 80 and 4, respectively. Pairwise surface areas were calculated similarly to Street and Mayo<sup>50</sup>. Separate scaling factors were stored for backbone and side chain atoms at each position. The scaling factors were selected so that the pairwise-calculated buried surface would equal the exact buried surface areas for all residues in the GCN4 structure (PDB entry 2ZTA)<sup>16</sup>. For sequences i–viii (Table 1), the average difference between pairwise computed and exact surface area was  $68 \text{ \AA}^2$ .  $U_{\text{LJ}}$  is the Lennard-Jones potential energy. For one-body energies, the Lennard-Jones function with OPLS-UA or CHARMM19 parameters was used. For interactions between rotamers (two-body energies), we used a fuzzy Lennard-Jones function. Lennard-Jones interaction energies were calculated with radii scaled<sup>51</sup> by 0.9, and negative (favorable) interaction energies were set to zero. Surface area buried between side chain rotamers was assigned an energy density of  $-16 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ ; this value was derived from two constants taken from the literature. First, the experimentally determined surface-area energy density for transfer of acetyl-X-amide analogs of non-polar side chains from water to octanol<sup>52,53</sup> is  $21 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ . Second, the FDPB /  $\gamma$  solvation model assigns a surface-area energy density for transfer of hydrocarbons from water to vacuum<sup>48</sup> of  $5 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ . The difference between these values,  $-16 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ , is the surface-area energy density for transfer from vacuum to octanol, the appropriate value for the fuzzy Lennard-Jones function.



**Design targets and competitors.** The eight designed positions in each calculation are located in two distinct polypeptides, A and B (Fig. 1). Free energies for the coiled-coil states were calculated using the coiled-coil backbone template subject to mean-field repacking of side chains. The free energy of the unfolded state was calculated in two steps according to the following scheme:  $D \leftrightarrow 2M \leftrightarrow 2U$ , where D is the folded dimer, M is a monomer  $\alpha$ -helix and U is the unfolded polypeptide monomer. Free energies for monomer helices were calculated using an isolated helix backbone subject to mean-field repacking of side chains. The free energy for unfolding of the monomer helices was computed using the AGADIR parameters<sup>54</sup>. We added a sequence-independent constant to the energy of the unfolded state so that the stability of the pCAP sequence would evaluate to 3 kcal mol<sup>-1</sup>, its measured stability at 1  $\mu$ M concentration. The aggregated state was modeled as the target conformation embedded in a medium with a dielectric constant of 65 rather than 80. This value yields an energy gap between the native and aggregated states of the pCAP sequence comparable to its stability at 1  $\mu$ M concentration. Inclusion of the aggregated state as a competitor guarantees that the designed sequences will have an unfavorable transfer free energy to a solvent of lower dielectric constant. The free energies of states involving heterodimers were decreased by  $RT \ln(2)$  to account for the entropy of mixing.

Each state in the design consisted of two copies of the A and B polypeptides in different environments and arrangements. The homo- and heterodimer states consisted of the appropriate arrangements of the A and B sequences evaluated in the folded conformation. Three unfolded-state competitors were considered, corresponding to the unfolding of AA, BB or one copy of AB, with the other polypeptides remaining folded. Designs targeting the homodimer state included two aggregated state competitors, and those targeting the heterodimer state included one. These competitors involved the transfer of single coiled coils to the lower dielectric environment, analogous to the unfolded-state competitors.

**Fitness function.** The fitness of a given sequence is defined as the transfer free energy of that sequence from the target state to the ensemble of competing states (Fig. 2b).

$$\text{Fitness} = -RT \ln(\sum_{\text{competitors}} e^{-A_c / RT}) - A_{\text{target}} \quad (3)$$

where  $A_c$  is the free energy of the competing states,  $A_{\text{target}}$  is the free energy of the target state and  $RT$  is evaluated at room temperature.

**Genetic algorithm.** An initially random population of 4,800 discrete sequences was propagated for 30 generations. Three rules dictated the composition of subsequent generations once fitness scores were evaluated. First, the most-fit sequence of each generation was automatically propagated. Uniform crossover recombination was then used to generate 99% of the remaining sequences<sup>55</sup>. Finally, mutation of single sequences at 20% probability per site was used to generate the remainder of the population. Sequences chosen for the recombination and mutation processes were selected randomly, biased by fitness scores such that  $P_s$  (the probability of selecting sequence  $s^*$ ) was

$$P_s = (e^{-F_s / \sigma}) / (\sum_s e^{-F_s / \sigma}) \quad (4)$$

where  $F_s$  is the fitness of member  $s$ ,  $\sigma$  is the standard deviation in fitness for the current generation and the sum in the denominator extends over the entire population. We performed each design calculation three times with different random initial sequence populations. These calculations identified the same

best sequence, suggesting that the genetic algorithm finds the global optimum.

**Cloning and expression.** The pCAP and pCAP-SH-PKA constructs (Fig. 1b) were appended to the *TrpLE'* leader sequence<sup>56</sup>. All constructs were cloned into the pET24a vector (Novagen) using standard molecular biology techniques. Mutations were introduced using the method of Kunkel<sup>57</sup> and verified by DNA sequencing. The pCAP and pCAP-SH-PKA peptides were purified from inclusion bodies and cleaved from the *TrpLE'* leader sequence with cyanogen bromide. Peptides in the pH6-CAP-SH-PKA construct were purified by nickel-NTA affinity chromatography directly from cell lysates. Final purification of all peptides was performed by reversed-phase HPLC. Peptide identities were confirmed by electrospray mass spectrometry. Protein concentrations were determined using the method of Edelhösch<sup>58</sup>.

**Measurement of specificity.** Redox exchange reactions were performed at 10  $\mu$ M peptide concentration in 5 mM Tris-HCl, pH 9.0, 50  $\mu$ M  $\beta$ -mercaptoethanol and 100  $\mu$ M 2-hydroxyethyl disulfide<sup>18</sup>. The reactions were equilibrated overnight and quenched by the addition of iodoacetamide to 10 mM for 1 h. The peptides were labeled by incubation with 5 U protein kinase A (Sigma) at 37 °C for 3 h at final concentrations of 5  $\mu$ M peptide, 5 mM Tris-HCl, pH 9.0, 0.005% (v/v) Triton X-100, 40  $\mu$ M ATP and 10  $\mu$ M [ $\gamma$ -<sup>32</sup>P]ATP. The labeled mixture was analyzed by SDS-PAGE using the Tris-tricine system<sup>59</sup> in the absence of reducing agents. Gel bands were quantitated on a Phosphorimager (Molecular Dynamics).  $\Delta G_{\text{spec}}^X$  is defined as  $-RT \ln(K^X / K^{\text{pCAP}})$ , where  $K^X = [A^X B^X]^2 / ([A^X A^X][B^X B^X])$  and  $K^{\text{pCAP}}$  denotes the equilibrium constant for the pCAP sequence. Results from exchange reactions initiated from the pure homodimer and pure heterodimer forms of the protein agreed to within 0.1 kcal mol<sup>-1</sup> in all cases, indicating that equilibrium had been reached<sup>18</sup>.

**Measurement of stability.**  $\Delta \Delta G_{\text{unt}}$  is equal to the difference in stability ( $\Delta G_{\text{unt}}$ ) between each designed peptide and the original pCAP peptide.  $\Delta G_{\text{unt}}$  was determined by urea denaturation in 5 mM potassium phosphate, pH 7.1, monitored by circular dichroism spectroscopy at 222 nm and 4 °C on an Aviv DS-62A spectropolarimeter at 5  $\mu$ M peptide concentration. Data were converted to the fraction of dimer unfolded. Where a folded baseline was unavailable, data were collected at low concentrations of urea in the presence of 20% (v/v) trifluoroethanol (TFE), a potent helix-inducing solvent. A folded baseline was then extracted from this data. For well-folded species, 20% TFE did not affect the y-intercept or slope of the baseline (data not shown). Where an unfolded baseline was unavailable, a value of zero was assumed. A two-state bimolecular model was used to fit the homodimer data, and a two-state unimolecular model was used to fit the disulfide-bonded heterodimer data. Stabilities were measured for sequence vii A in both the disulfide-bonded and unmodified forms to serve as a calibration between the data sets.

#### Acknowledgments

We thank F.E. Boas, J.A. Silverman, D.R. Halpin, S.J. Wrenn and R.L. Baldwin for stimulating conversations and criticism during the course of this work and for comments on the manuscript. We also acknowledge helpful suggestions from the anonymous referees. This research was funded by a Searle scholar grant to P.B.H. from the Chicago Community Trust.

#### Competing interests statement

The authors declare that they have no competing financial interests.

Received 10 July, 2002; accepted 8 November, 2002.



1. Harbury, P.B., Plecs, J.J., Tidor, B., Alber, T. & Kim, P.S. High-resolution protein design with backbone freedom. *Science* **282**, 1462–1467 (1998).
2. Nauli, S., Kuhlman, B. & Baker, D. Computer-based redesign of a protein folding pathway. *Nat. Struct. Biol.* **8**, 602–605 (2001).
3. Dahiyat, B.I. & Mayo, S.L. De novo protein design: fully automated sequence selection. *Science* **278**, 82–87 (1997).
4. Pabo, C. Molecular technology. Designing proteins and peptides. *Nature* **301**, 200 (1983).
5. Ponder, J.W. & Richards, F.M. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **193**, 775–791 (1987).
6. Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. USA* **97**, 10383–10388 (2000).
7. Gordon, D.B., Marshall, S.A. & Mayo, S.L. Energy functions for protein design. *Curr. Opin. Struct. Biol.* **9**, 509–513 (1999).
8. Hellinga, H.W. & Richards, F.M. Optimal sequence selection in proteins of known structure by simulated evolution. *Proc. Natl. Acad. Sci. USA* **91**, 5803–5807 (1994).
9. Desjarlais, J.R. & Handel, T.M. De novo design of the hydrophobic cores of proteins. *Protein Sci.* **4**, 2006–2018 (1995).
10. Lazar, G.A., Desjarlais, J.R. & Handel, T.M. De novo design of the hydrophobic core of ubiquitin. *Protein Sci.* **6**, 1167–1178 (1997).
11. Malakauskas, S.M. & Mayo, S.L. Design, structure and stability of a hyperthermophilic protein variant. *Nat. Struct. Biol.* **5**, 470–475 (1998).
12. Marshall, S.A. & Mayo, S.L. Achieving stability and conformational specificity in designed proteins via binary patterning. *J. Mol. Biol.* **305**, 619–631 (2001).
13. Marvin, J.S. & Hellinga, H.W. Conversion of a maltose receptor into a zinc biosensor by computational design. *Proc. Natl. Acad. Sci. USA* **98**, 4955–4960 (2001).
14. Wilson, C., Mace, J.E. & Agard, D.A. Computational method for the design of enzymes with altered substrate specificity. *J. Mol. Biol.* **220**, 495–506 (1991).
15. Shimaoka, M. et al. Computational design of an integrin I domain stabilized in the open high affinity conformation. *Nat. Struct. Biol.* **7**, 674–678 (2000).
16. O'Shea, E.K., Klemm, J.D., Kim, P.S. & Alber, T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **254**, 539–544 (1991).
17. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
18. O'Shea, E.K., Rutkowski, R., Stafford, W.F. III & Kim, P.S. Preferential heterodimer formation by isolated leucine zippers from Fos and Jun. *Science* **245**, 646–648 (1989).
19. O'Shea, E.K., Lumb, K.J. & Kim, P.S. Peptide Velcro: design of a heterodimeric coiled coil. *Curr. Biol.* **3**, 658–667 (1993).
20. Jorgensen, W.L. & Tiradadorives, J. The OPLS potential functions for proteins: energy minimizations for crystals of cyclic-peptides and crambin. *J. Am. Chem. Soc.* **110**, 1666–1671 (1988).
21. Brooks, B.R. et al. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
22. Koehl, P. & Levitt, M. De novo protein design. II. Plasticity in sequence space. *J. Mol. Biol.* **293**, 1183–1193 (1999).
23. Janin, J. Quantifying biological specificity: the statistical mechanics of molecular recognition. *Proteins* **25**, 438–445 (1996).
24. Ghirlanda, G., Lear, J.D., Lombardi, A. & DeGrado, W.F. From synthetic coiled coils to functional proteins: automated design of a receptor for the calmodulin-binding domain of calcineurin. *J. Mol. Biol.* **281**, 379–391 (1998).
25. Hellinga, H.W. Rational protein design: combining theory and experiment. *Proc. Natl. Acad. Sci. USA* **94**, 10015–10017 (1997).
26. Hellinga, H.W. Construction of a blue copper analogue through iterative rational protein design cycles demonstrates principles of molecular recognition in metal center formation. *J. Am. Chem. Soc.* **120**, 10055–10066 (1998).
27. Yue, K. et al. A test of lattice protein-folding algorithms. *Proc. Natl. Acad. Sci. USA* **92**, 325–329 (1995).
28. Deutsch, J.M. & Kurosky, T. New algorithm for protein design. *Phys. Rev. Lett.* **76**, 323–326 (1996).
29. Irback, A., Peterson, C., Potthast, F. & Sandelin, E. Monte Carlo procedure for protein design. *Phys. Rev. E* **58**, R5249–R5252 (1998).
30. Dima, R.I., Banavar, J.R., Cieplak, M. & Maritan, A. Statistical mechanics of protein-like heteropolymers. *Proc. Natl. Acad. Sci. USA* **96**, 4904–4907 (1999).
31. Banavar, J.R. et al. Structure-based design of model proteins. *Proteins* **31**, 10–20 (1998).
32. Rossi, A., Maritan, A. & Micheletti, C. A novel iterative strategy for protein design. *J. Chem. Phys.* **112**, 2050–2055 (2000).
33. Koehl, P. & Levitt, M. De novo protein design. I. In search of stability and specificity. *J. Mol. Biol.* **293**, 1161–1181 (1999).
34. Jones, S. & Thornton, J.M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA* **93**, 13–20 (1996).
35. Buckle, A.M., Schreiber, G. & Fersht, A.R. Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-Å resolution. *Biochemistry* **33**, 8878–8889 (1994).
36. Shoichet, B.K., Baase, W.A., Kuroki, R. & Matthews, B.W. A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci. USA* **92**, 452–456 (1995).
37. Derewenda, U. et al. An unusual buried polar cluster in a family of fungal lipases. *Nat. Struct. Biol.* **1**, 36–47 (1994).
38. Warshel, A. & Aqvist, J. Electrostatic energy and macromolecular function. *Annu. Rev. Biophys. Chem.* **20**, 267–298 (1991).
39. Lumb, K.J. & Kim, P.S. A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry* **34**, 8642–8648 (1995).
40. Bolon, D.N. & Mayo, S.L. Polar residues in the protein core of *Escherichia coli* thioredoxin are important for fold specificity. *Biochemistry* **40**, 10047–10053 (2001).
41. Desjarlais, J.R. & Handel, T.M. Side-chain and backbone flexibility in protein core design. *J. Mol. Biol.* **290**, 305–318 (1999).
42. Keating, A.E., Malashkevich, V.N., Tidor, B. & Kim, P.S. Side-chain repacking calculations for predicting structures and stabilities of heterodimeric coiled coils. *Proc. Natl. Acad. Sci. USA* **98**, 14825–14830 (2001).
43. Jencks, W.P. *Catalysis in Chemistry and Enzymology* (Dover, New York; 1987).
44. Harbury, P.B., Tidor, B. & Kim, P.S. Repacking protein cores with backbone freedom: structure prediction for coiled coils. *Proc. Natl. Acad. Sci. USA* **92**, 8408–8412 (1995).
45. Dunbrack, R.L. & Karplus, M. Backbone-dependent rotamer library for proteins. *J. Mol. Biol.* **230**, 543–574 (1993).
46. Koehl, P. & Delarue, M. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J. Mol. Biol.* **239**, 249–275 (1994).
47. Mendes, J., Soares, C.M. & Carrondo, M.A. Improvement of side-chain modeling in proteins with the self-consistent mean field theory method based on an analysis of the factors influencing prediction. *Biopolymers* **50**, 111–131 (1999).
48. Sitkoff, D., Sharp, K.A. & Honig, B. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **98**, 1978–1988 (1994).
49. Havranek, J.J. & Harbury, P.B. Tanford-Kirkwood electrostatics for protein modeling. *Proc. Natl. Acad. Sci. USA* **96**, 11145–11150 (1999).
50. Street, A.G. & Mayo, S.L. Pairwise calculation of protein solvent-accessible surface areas. *Fold. Des.* **3**, 253–258 (1998).
51. Dahiyat, B.I. & Mayo, S.L. Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci. USA* **94**, 10172–10177 (1997).
52. Fauchère, J.-L. & Pliska, V. Hydrophobic parameters- $\pi$  of amino-acid side-chains from the partitioning of N-acetyl-amino-acid amides. *Eur. J. Med. Chem.* **18**, 369–375 (1983).
53. Wimley, W.C., Creamer, T.P. & White, S.H. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35**, 5109–5124 (1996).
54. Lacroix, E., Viguera, A.R. & Serrano, L. Elucidating the folding problem of  $\alpha$ -helices: local motifs; long-range electrostatics; ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.* **284**, 173–191 (1998).
55. Mitchell, M. *An Introduction to Genetic Algorithms* (MIT Press, Cambridge, Massachusetts; 1996).
56. Kleid, D.G. et al. Cloned viral protein vaccine for foot-and-mouth disease: responses in cattle and swine. *Science* **214**, 1125–1129 (1991).
57. Kunkel, T.A., Bebenek, K. & McClary, J. Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.* **204**, 125–139 (1991).
58. Edelhoch, H. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**, 1948–1954 (1967).
59. Schagger, H. & von Jagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166**, 368–379 (1987).
60. Lu, M. et al. Helix capping in the GCN4 leucine zipper. *J. Mol. Biol.* **288**, 743–752 (1999).
61. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).



## Human Biliverdin Reductase Is a Leucine Zipper-like DNA-binding Protein and Functions in Transcriptional Activation of Heme Oxygenase-1 by Oxidative Stress\*

Received for publication, August 27, 2001, and in revised form, November 5, 2001  
Published, JBC Papers in Press, December 31, 2001, DOI 10.1074/jbc.M108239200

Zulfiqar Ahmad†, Mohammad Salim, and Mahin D. Maines§

From the Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642

Human biliverdin reductase (hBVR) is a serine/threonine kinase that catalyzes reduction of the heme oxygenase (HO) activity product, biliverdin, to bilirubin. A domain of biliverdin reductase (BVR) has primary structural features that resemble leucine zipper proteins. A heptad repeat of five leucines (L<sub>1</sub>–L<sub>5</sub>), a basic domain, and a conserved alanine characterize the domain. In hBVR, a lysine replaces L<sub>3</sub>. The secondary structure model of hBVR predicts an  $\alpha$ -helix-turn- $\beta$ -sheet for this domain. hBVR translated by the rabbit reticulocyte lysate system appears on a nondenaturing gel as a single band with molecular mass of ~69 kDa. The protein on a denaturing gel separates into two anti-hBVR immunoreactive proteins of ~39.9 + 34.6 kDa. The dimeric form, but not purified hBVR, binds to a 100-mer DNA fragment corresponding to the mouse HO-1 (hsp32) promoter region encompassing two activator protein (AP-1) sites. The specificity of DNA binding is suggested by the following: (a) hBVR does not bind to the same DNA fragment with one or zero AP-1 sites; (b) a 56-bp random DNA with one AP-1 site does not form a complex with hBVR; (c) *in vitro* translated HO-1 does not interact with the 100-mer DNA fragment with two AP-1 sites; (d) mutation of Lys<sup>143</sup>, Leu<sup>150</sup>, or Leu<sup>157</sup> blocks both the formation of the ~69-kDa specimens and hBVR DNA complex formation; and (e) purified preparations of hBVR or hHO-1 do not bind to DNA with two AP-1 sites. The potential significance of the AP-1 binding is suggested by the finding that the response of HO-1, in COS cells stably transfected with antisense hBVR, with 66% reduced BVR activity, to superoxide anion (O<sub>2</sub><sup>-</sup>) formed by menadione is attenuated, whereas induction by heme is not affected. We propose a role for BVR in the signaling cascade for AP-1 complex activation necessary for HO-1 oxidative stress response.

Biliverdin reductase (BVR)<sup>1</sup> is a recently described serine/threonine kinase (1) that catalyzes reduction of biliverdin

IX $\alpha$  at the  $\gamma$  meso bridge to produce bilirubin. Biliverdin is the product of heme (Fe-protoporphyrin IX) oxidation by the heme oxygenase (HO) system. The reductase in response to extracellular stimuli (e.g. cGMP, lipopolysaccharides, and free radicals) translocates into the nucleus (2) and is activated by oxygen radicals (1). The mammalian enzyme is highly conserved; the rat and human reductases share 84% amino acid identity (3, 4). Certain features of the reductase are conserved phylogenetically from cyanobacteria to humans including its unique property among all enzymes characterized to date of having dual pH/cofactor requirement (5, 6). Human BVR (hBVR) is a 296-residue-long polypeptide that, based on its predicted amino acid sequence, has a region with certain key residues that are conserved in proteins that have a leucine zipper dimerization domain, such as human Shaker, human c-Myc, *Saccharomyces* GCN4, human c-Jun, human CREB, human c-Fos, and *Saccharomyces* YAP-1 (Fig. 1). This motif is also found in the rat enzyme (Fig. 1). As a rule, the leucine zipper motif consists of repeat of five leucines (L<sub>1</sub>–L<sub>5</sub>) separated by six amino acids (Fig. 1) (7, 8). Exceptions to this, however, are found, for instance in *Saccharomyces* YAP-1: L<sub>3</sub> is substituted with asparagine; in *Saccharomyces* GCN4 and human CREB, L<sub>5</sub> is substituted by arginine and lysine, respectively; and, in human c-Myc, valine replaces L<sub>1</sub>. They all form functional homodimers or heterodimers. In hBVR and rat BVR, L<sub>3</sub> is substituted with lysine at positions 143 and 142, respectively (Fig. 1). Other structural features of the dimerization domain include a secondary structure that in most cases fits the helix-turn-helix model (8–10) and an invariant basic region that starts exactly seven residues N-terminal to L<sub>1</sub> and is flanked by alanine residues (Fig. 1). The basic region is the DNA binding domain (7, 8, 11). An  $\alpha/\beta$  secondary structure with leucine-rich repeats also forms a high affinity protein-protein interaction domain (12, 13). Although the leucine zipper dimerization motif has been identified in several nonnuclear proteins (Fig. 1), the greater numbers of proteins that have these conserved features are transacting factors and play a role in regulation of gene expression.

The AP-1 site is one of the DNA recognition sequences for leucine zipper proteins. The heme oxygenase cognate, HO-1 or hsp32 (14) is activated by increased AP-1 DNA binding in response to certain oxidative stress stimuli (15, 16). Transcriptional activation involves binding of c-Jun and c-Fos homodimers or heterodimers to the AP-1 site (17, 18). Increased AP-1 complex formation is not restricted to HO-1 or oxidative stress; rather, it is identified for activation of several oncogenes and kinases in response to cytokines, growth factors, transformation factors, UV radiation, and other assorted stimuli (19).

\* This work was supported by National Institutes of Health Grants ES04066 and ES04391. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† On leave from the Department of Biochemistry, Hamdard University, Hamdard Nagar, New Delhi 10062, India.

§ To whom correspondence should be addressed: Dept. of Biochemistry/Biophysics, University of Rochester Medical Center, Box 712, 601 Elmwood Ave., Rochester, NY 14642. Tel.: 716-275-5383; Fax: 716-275-6007; E-mail: mahin\_maines@urmc.rochester.edu.

<sup>1</sup> The abbreviations used are: BVR, biliverdin reductase; HO, heme oxygenase; hBVR, human BVR; AP-1, activator protein; MD, menadione.

Using the x-ray diffraction analysis of rat BVR (20)<sup>2</sup> and alignment of the predicted amino acid sequence of hBVR (4), we have identified conserved features of leucine zipper DNA-binding proteins in the reductase. We have questioned whether hBVR recognizes specific sequences of DNA and, if so, whether this binding is of biological significance. We present data that show specific binding of native hBVR to DNA and suggest a role for BVR in regulation of HO-1 oxidative stress response.

## EXPERIMENTAL PROCEDURES

### Materials

All of the chemicals and biochemicals used in this study were of ultrapure quality purchased from Sigma, Aldrich, or Invitrogen. Enzymes used in this study (*Bam*HI, *Blp*I, *Hind*III, *Sal*I, *Sma*I, *Xho*I, T4 DNA ligase, DNA polymerase, and polynucleotide kinase) were purchased from New England Biolabs, Invitrogen, or Amersham Biosciences, Inc. [<sup>35</sup>S]methionine and [<sup>32</sup>P]ATP Redivue<sup>TM</sup> radioisotopes were purchased from Amersham Biosciences. We used Redivue<sup>TM</sup> L-[<sup>35</sup>S]methionine (catalog no. AG 1094), because this grade of [<sup>35</sup>S]methionine does not cause the background labeling of the rabbit reticulocyte lysate 42-kDa protein that can occur using other grades of labels (21).

### Methods

**In Vitro Synthesis of Capped RNA Transcript**—The full-length BVR fragment was amplified from the plasmid 494 Gex3 (4) using oligonucleotides OL.507 and OL.508, while HO-1 (22) was amplified using oligonucleotides OL.547 and OL.548 (Table I). They were inserted in the multiple cloning site of pCDNA3 (Invitrogen) between *Bam*HI and *Xho*I. The resultant recombinant DNAs were named as p507 and p547. Methods used in the construction of plasmids, including restriction enzyme digestion, separation of plasmid DNA and restriction fragments on agarose gels, ligation of DNA fragments, and the isolation of plasmid DNA are described in Sambrook *et al.* (23). *Escherichia coli* transformations were performed with CaCl<sub>2</sub> (24). PCR was carried out as described by Saiki *et al.* (25). Both plasmid p507 and p547 were transformed in INV-competent cells. The plasmid purification was done with Qiagen Mini Prep plasmid purification kit and was linearized by digesting with *Sma*I. Linearized plasmid was then treated with phenol/chloroform/isoamyl alcohol (25:24:1) and ethanol-precipitated. Plasmids were dissolved and stored in RNase-free water. RNA was transcribed by using the RiboProbe *in vitro* Transcription System from Promega. 5 µg of linearized template DNA was used in a 50-µl reaction volume using T7 RNA polymerase in the presence of the m7G cap analog so as to generate the capped transcript. 50 units of ribonuclease inhibitor were also added to the reaction along with required amounts of dithiothreitol and nucleotides. After a 1-h incubation at 37 °C, the reaction mixture was treated with RNase-free DNase (1 µl/µg of template DNA) and was extracted with phenol/chloroform/isoamyl alcohol, precipitated with ethanol and ammonium acetate, and resuspended in 20 µl of RNase-free water and kept at -70 °C.

**In Vitro Translation**—A 5.4-kb pcDNA 3 with 1 kb coding hBVR was used as vector to generate *in vitro* transcribed mRNA with T7 RNA polymerase. The transcribed mRNA was translated in the presence of [<sup>35</sup>S]methionine using rabbit reticulocyte lysate. *In vitro* translation was performed using micrococcal nuclease-treated rabbit reticulocyte lysate (Promega). A 50-µl reaction mixture was prepared by using 35 µl of lysate, 1 µl of 0.1 M dithiothreitol, 2 µl of 1 mM amino acid mixture minus methionine, 1 µl of RNase inhibitor and 5 µl of translation grade [<sup>35</sup>S]methionine. 5 µl of transcribed mRNA was added to the above reaction mixture and immediately incubated at 30 °C for 90 min. The *in vitro* translated proteins were resolved on 12% SDS or native polyacrylamide gel along with rainbow or native high molecular weight markers, respectively (Amersham Pharmacia Biotech). The gels were fixed in 10% acetic acid and 30% methanol and then treated with autoradiography enhancer (Amplify; Amersham Biosciences) for 30 min and dried under vacuum at 80 °C for 2 h and autoradiographed at -70 °C.

**Preparation of <sup>32</sup>P-labeled DNA Fragments**—A 56- or 100-bp DNA fragment with and without AP-1 sites was used for the DNA binding assay; their sequences are shown in Table I (OL.619, OL.620; OL.623–OL.630). Complementary oligonucleotides were used to generate double-stranded DNA fragments. 150-ng aliquots of annealed oligonucleo-

tides were radioactively labeled using [<sup>32</sup>P]ATP and T4 polynucleotide kinase. The DNA probes were purified with the Qiagen Nucleic Acid Purification Kit.

**PCR-generated Site-directed Mutagenesis**—A 1-kb hBVR fragment was cut out from plasmid p507 by *Sal*I. This 1-kb fragment was used as the template DNA for site-directed mutagenesis. Oligonucleotides (OL.582–OL.587) used for mutagenesis of hBVR leucine zipper motif at positions Lys<sup>143</sup>, Leu<sup>150</sup>, and Leu<sup>157</sup> are shown in Table I. PCR was carried out in two steps. In the first step, the substitutions were introduced by using OL.621 or OL.622 in combination with oligonucleotides OL.582 and OL.583, OL.584 and OL.585, or OL.586 and OL.587 to generate K143A, L150A, and L157A, respectively. In the second stage of the reaction, the PCR products from the first stage were used as template DNA and were joined together by using oligonucleotides OL.621 and OL.622 (Table I). Another difference in the two-step 30-cycle PCRs was the *T<sub>m</sub>*, which was 48 °C in the first reaction and 43 °C in the second. The PCR products, thus formed, were purified with PCR purification kit (Concert) and digested with *Blp*I and *Hind*III. The resultant fragments were inserted in p507, which was used as a vector. Ligation was done within the gel by using 1% low melt agarose. The plasmids were amplified in XL-1 Blue cells and isolated by the Qiagen Mini Prep kit. The DNA sequencing of the mutated hBVR segment was carried out with the oligonucleotides OL.582–OL.587 (Table I) using the ABI PRISM dye Terminator Cycle Sequencing Ready Reaction kit with AmpliTaq DNA polymerase (Big Dye).

**Native and Denaturing Gel Analyses**—*In vitro* translated protein was assayed on native gel immediately after synthesis. One µl of *in vitro* translated material was added to 2 µl (25 ng) of annealed, unlabeled control DNA fragment. To this, 0.4 µg of poly(dI-dC) (Amersham Biosciences) in 14 µl of DNA binding buffer (10 mM Tris-chloride (pH 7.4), 50 mM NaCl, 1 mM MgCl<sub>2</sub>, 1 mM EDTA, 1 mM dithiothreitol, 5% glycerol) was added. It was incubated for 5 min at room temperature, and after adding 5 µl of loading buffer (1.5× DNA binding buffer with bromophenol blue dye), samples were resolved on 12% native polyacrylamide gel in Tris-acetate/EDTA buffer at 35 milliamperes. The control DNA helps to prevent the formation of nonspecific protein aggregates, thereby increasing the resolution of protein bands (26). A portion of the translated protein was treated with SDS and analyzed on a denaturing 12% polyacrylamide gel.

**DNA Binding Assay**—As with native gel analysis, *in vitro* translated proteins were assayed for DNA binding immediately after synthesis. 1 µl of translated material was added to 5000–500,000 cpm of <sup>32</sup>P-labeled DNA fragment representing ~2–3 ng of DNA. 0.1 µg of poly(dI-dC) in 10 µl of DNA binding buffer was added to the labeled DNA. After incubating samples for 20 min at room temperature, 5 µl of loading buffer was added. The samples were resolved on 12% native polyacrylamide gel with 35 milliamperes at 4 °C. The gels were processed as described above. Dried gels were put on two pieces of film separated by a piece of paper. Autoradiography was done at -70 °C for different time periods.

**Western Blot Analysis**—For Western blot analysis, the primary antibody was rabbit anti-human kidney BVR (27) with ECL detection system RPN 2106 (Amersham Biosciences). Briefly, *in vitro* translated hBVR was subjected to 12% SDS-polyacrylamide gel, transferred to polyvinylidene difluoride transfer membrane (Pall Corp.), and subjected to Western blot analysis as described earlier (1).

**COS Cell Transfection and BVR Measurement**—A cytotoxicity curve for the drug G418 sulfate (Geneticin), used as a marker for the selection of clonal cell lines, was established for exponentially grown COS cells in Dulbecco's modified Eagle's medium (37 °C, 5% CO<sub>2</sub>). At a concentration of 440 mg/ml and beyond, the drug was found toxic to the parental cell line. Therefore, the selection medium contained G418 at a concentration of 450 mg/ml. pcDNA3 plasmid containing the antisense sequence was isolated from *E. coli* cultures using Qiagen Midi Prep kit. Transfection was carried out by electroporation. The following day, transfected cells were split 1:2 and seeded on a 100-mm culture dish in the selection medium. The selection process was continued for 8–10 days with a change of selection medium every 2 days. Cells grown in culture flasks to 75% confluence were pooled from three flasks and were used for BVR enzyme activity measurement and mRNA analysis. BVR activity was measured from an increase in absorbance at 450 nm as described before (5) using bilirubin as the substrate and NADH as the cofactor. The activity is expressed as units, a unit representing 1 nmol of bilirubin formed/min/mg of protein.

**Northern Blotting**—The HO-1 hybridization probe was a 569-base pair HO-1 fragment corresponding to nucleotides 86–654 of rat HO-1 cDNA (28). Cells from a minimum of three culture flasks were pooled and used for each analysis. Total RNA was extracted from COS cells for preparation of poly(A)<sup>+</sup> RNA that was separated by electrophoresis on

<sup>2</sup> F. Whitby, J. Phillips, W. K. McCoubrey, C. Hills, and M. D. Maines, unpublished results.

		BASIC DOMAIN	LEUCINE ZIPPER	
hBVR	100	MTLSLAAQQLMELDEOKGKVLHEEHVELLMEEFAFLKKEVVGKDLLKGSLLFTSDPL	1 2 3 4 5	157
rBVR	99	MTLSFAAAQQLMELAAOKGRVLEEHVELLMEEFEPLRRVVGKDLLKGSLLFTASPL		156
hSHAKER	326	GGGQNGQOQMSLAIRVILRVFRIPKLSRHSKGLQIGKTKLQASMRREGLLIFPL		383
hC-MYC	377	LRDQIPELENNEKAPKVILKATAYILSVQAEQFLISEEDLLKRRRQQLKHLEQL		434
sGCN4	224	SSDPAALKRRARNTAAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKLVGSR		281
hC-JUN	251	ERIKAEKRMENRIAAASKCRKKLERIARLEEKVKTLKAQNSLASTANMLREQVAQL		308
hCREB	282	AARKREVLKMKNRDAARECRKKKEYVVKLENRAVAVLENQNTLIEELKALKDLYCHK		339
hC-FOS	136	EEKRRIRRRRNMAAAKCNRRRELDTTQAEEDQLEDEKSAQTETANLLEKEKL		193
sYAP-1	63	DPETKQKRTAQNRRAAFRRERKERKMKLEKKVQSLSEICQONEVEATFLRQDLITL		120

FIG. 1. Amino acid alignment of leucine zipper protein domains. Key leucine zipper domain molecules ( $L_1$ – $L_5$ ) and their respective replacements are shown in **boldface type**. Human Shaker, c-Jun, and c-Fos have all five ( $L_1$ – $L_5$ ) leucine molecules, whereas in the case of hBVR, rat BVR, human c-Myc, *Saccharomyces* GCN4, human CREB, and *Saccharomyces* YAP-1 leucine molecules at positions  $L_3$ ,  $L_4$ ,  $L_5$ ,  $L_6$ , and  $L_7$  are substituted with lysine, lysine, valine, arginine, lysine, and asparagine, respectively. The basic domain is shown as cluster-spacer-cluster structure, and the basic residues are underlined. Sequences are derived from *Homo sapiens* (h), *Rattus norvegicus* (r), and *Saccharomyces cerevisiae* (s) (14, 47–53).

denaturing formaldehyde gel, and transferred onto a Nytran membrane. The HO-1 and actin probes were labeled using [ $\alpha$ - $^{32}$ P]dCTP with the Random Primers Labeling System (Invitrogen). Prehybridization and hybridization were performed as described previously (29). Blots were probed sequentially with HO-1 and actin. The signals were quantitated using TempDens Platform version 1.0.0 and are expressed relative to that of the control. The control level is arbitrarily given the value of 1.

## RESULTS

The comparison of the primary structure of hBVR between amino acids 100 and 157 with known leucine zipper-type DNA binding proteins shows certain common features (Fig. 1). These include the five repeating amino acids  $L_1$ ,  $L_2$ ,  $K_3$ ,  $L_4$ , and  $L_5$ , spaced every seventh residue, and a basic domain that is flanked by an upstream alanine residue and starts exactly seven residues N-terminal to  $L_1$ . There are, however, differences in the primary structure of hBVR and those of most leucine zipper DNA binding proteins; a second basic domain that is present in DNA-binding proteins GCN4, c-Jun, c-Fos, and YAP-1 is not present in BVR. Fig. 2 shows the secondary structure of hBVR, which is modeled after x-ray diffraction analyses of rat BVR crystal structure and shows a U-shaped  $\alpha$ -helix-turn- $\beta$  motif for the leucine zipper motif. Residues that form heptads are identified by a space-filling model. It is noted that a leucine-rich  $\alpha$ -helix-turn- $\beta$  structure is also present in porcine ribonuclease inhibitor and is involved in heterodimer and homodimer formations (12, 13). On the basis of the crystal structure, Kobe and Deisenhofer (12, 13) have shown that the leucine-rich repeat of the ribonuclease inhibitor is also "horseshoe-shaped."

**hBVR Forms a Homodimer and Binds DNA**—Observations with the primary and secondary features of hBVR were followed by examination of whether hBVR forms a dimer, and if so, whether the dimer interacts with DNA. For DNA interaction analysis, 56-mer and a 100-mer (Table I) DNA fragments encompassing AP-1 sites were used. The 56-mer fragment was a random fragment with one AP-1 site used for investigation of c-Jun and c-Fos DNA binding (26). AP-1 also has been tested for GCN4 binding (30). The 100-mer DNA fragment corresponded to the HO-1 promoter region encompassing two AP-1 sites (31). In order to bind to DNA, leucine zipper type proteins form a dimer, which takes place at the leucine zipper motif (32, 33). Most proteins bearing this structural feature form homodimers, and dimer formation is required for its efficient DNA binding. The only known exception, Fos, forms a stable

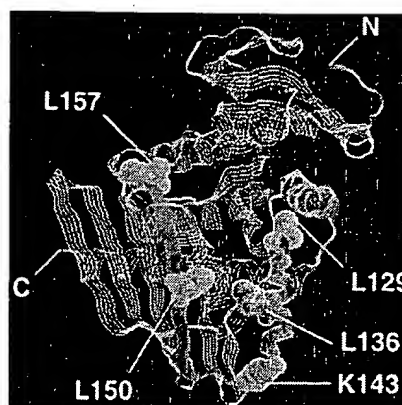


FIG. 2. The predicted three-dimensional structure of hBVR. Rat BVR coordinates were used to model the three-dimensional structure of hBVR. The residues of the leucine zipper (green and red) at key positions Leu<sup>129</sup>, Leu<sup>136</sup>, Lys<sup>143</sup>, Leu<sup>150</sup>, and Leu<sup>157</sup> are shown in the space-filling model. Residues between Leu<sup>129</sup> and Lys<sup>143</sup> are predicted to form an  $\alpha$ -helix; those between Lys<sup>143</sup> and Leu<sup>157</sup> form a  $\beta$ -sheet. N and C denote the N and C terminus, respectively. The figure was generated with the molecular graphic program RasMol (36).

heterodimer with Jun oncoprotein (17). Therefore, we examined hBVR for homodimer formation immediately after *in vitro* translation of hBVR mRNA, using cold native polyacrylamide gel (4 °C), and employed denaturing/SDS-polyacrylamide gel to dissociate the dimer immediately after *in vitro* translation of hBVR mRNA, should it be formed. On the native gel, the translated protein migrated as an approximately 69-kDa protein (Fig. 3). The protein size was assessed using standard native high molecular weight markers (Amersham Biosciences). Nonspecific protein aggregation was prevented by the addition of control unlabeled DNA (26).

Next, whether the protein synthesized by reticulocyte lysate is in fact hBVR was tested. For this, the *in vitro* translated protein was examined on a 12% SDS-polyacrylamide gel, and the gel was processed either for autoradiography (Fig. 4A) or for Western blot analysis (Fig. 4B). As shown in the autoradiogram, two prominent bands at ~35 and ~40 kDa were detected. hBVR, based on its predicted amino acid composition, has a molecular mass of ~34 kDa (4). However, because of extensive posttranslational modification, it migrates as a group of size variants with an approximate molecular mass in the range of ~38–42 kDa in SDS gel (4, 34). The Western blot shows, when *in vitro* translated hBVR is probed with antibody to human kidney BVR, two closely migrating bands. The identity of the translated protein was confirmed by comparing its gel migration with wild type hBVR and comparing its immunoreactivity with antibody with that of purified human kidney BVR. As noted in Fig. 4B, the pattern of immunostaining of proteins was nearly identical. The control consisted of the rabbit reticulocyte lysate without the addition of transcribed hBVR mRNA. In this lysate, bands near hBVR antibody-immunoreactive bands at the 35–40-kDa region were not detected. Collectively, these findings suggested that hBVR is capable of forming a homodimer.

To determine whether the synthesized hBVR binds to DNA, the *in vitro* translated hBVR was incubated with  $^{32}$ P-labeled 56-mer or 100-mer DNA fragments. An identical 56-bp fragment in which the AP-1 site was substituted with an unrelated sequence of equal length was used as control DNA. In addition, two identical 100-bp fragments with one AP-1 or zero AP-1 sites were synthesized and used as controls (OL.619, OL.620; OL.623–OL.630; Table I). After translation, the protein was incubated with DNA fragment, and the protein/DNA mixture was run on a native nondenaturing polyacrylamide gel. To

TABLE I  
List of oligonucleotides

The substitutions K143A, L150A, and L157A in oligonucleotides OL.582–OL.587 are shown in boldface type and are underlined. The AP-1 sites are also shown in boldface type and are underlined, while the replacements of AP-1 sites by random sequences are shown in boldface only for the oligonucleotides OL.619, OL.620, and OL.623–OL.630.

Oligonucleotide number	Sequence
OL.507	GGATCCATGAATGCAGAGCCCGAGAG
OL.508	CTCGAGAGCTACATCACCTCCTCCTC
OL.547	GGATCCATGGAGCGCCACAGCTCG
OL.548	GCTCGAGTGGCGAAGGATCACCATCGCAGGAGCGGTGT
OL.582	GAAGAAAGAAAGTGGTGGGGGCTGACCTGCTGAAAGGGTCTG
OL.583	CGACCCCTTCAGCAGGTCAGCCCCACCACTCTTTTTC
OL.584	GACCTGCTGAAAGGGTCTGCGCCTCTTCACATCTGACCCG
OL.585	CGGGTCAGATGTGAAGAGGGCCGACCCCTTCAGCAGGTC
OL.586	CCTCTTCACATCTGACCCGGCTGAAGAAGACCGGTTGGCT
OL.587	AGCCAAACCGGTCTTCTTCAGCCGGGTCAGATGTGAAGAGG
OL.619	TCCTCAGCTGCTTTTATGCT <b>TGTGTCAT</b> TGGTTGGGAGGGGTGATTAGCAGACAAAGGAAGACAGATTTTGCGATCCTCCC
OL.620	CTCTGTTCCCTCTGCCTCAG CTGAGGCAGAGGGAACAGAGGGGAGGATCGCAAAATCTGTCTTCCCTTTGTCTGCTAATCACCCCTCCCAACCATA <b>CACACA</b> GCATAAAAGCAGCTGAGGA
OL.621	CAGCCATGAGGACTACATCAG
OL.622	AGCCAGTTCCTTCTCAGAGAA
OL.623	TCCTCAGCTGCTTTTATGCT <b>TGTGTCAT</b> TGGTTGGGAGGGGTGATTAGCAGACAAAGGAAGACAGATTTTGCT <b>GAGTCACC</b> CTCTGTTCCCTCTGCCTCAG
OL.624	CTGAGGCAGAGGGAACAGAGGGT <b>GACTCAG</b> CACAAATCTGTCTTCCCTTTGTCTGCTAATCACCCCTCCCAACCAT <b>GACACA</b> GCATAAAAGCAGCTGAGGA
OL.625	TCCTCAGCTGCTTTTATGCGATCCTCTGGTTGGGAGGGGTGATTAGCAGACAAAGGAAGACAGATTTTGCGATCCTCCC CTCTGTTCCCTCTGCCTCAG
OL.626	CTGAGGCAGAGGGAACAGAGGGGAGGATCGCAAAATCTGTCTTCCCTTTGTCTGCTAATCACCCCTCCCAACCAGAGGAT CGCATAAAAGCAGCTGAGGA
OL.627	CACTGAGAGAACTATTACACAAGCCACATTAGCAT <b>GACTCATT</b> GTTTCTGATCAG
OL.628	CTGATCAGAAACAAT <b>GAGTCAT</b> GCTAATGTGGCTTGTGTAATAGTTTCTCTCAGTG
OL.629	CACTGAGAGAACTATTACACAAGCCACATTAGCAGATCCTCTTGTGTTTCTGATCAG
OL.630	CTGATCAGAAACAAGAGGATCTGCTAATGTGGCTTGTGTAATAGTTTCTCTCAGTG

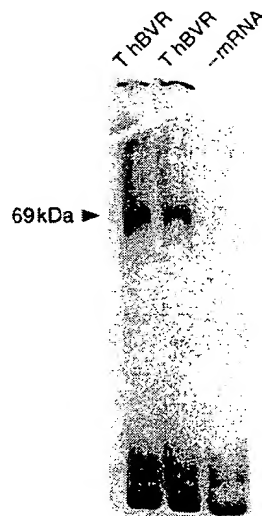


FIG. 3. Detection of high molecular weight protein synthesized by hBVR mRNA. A, *in vitro* translated hBVR as visualized on a 12% native polyacrylamide gel. From the left, the first two lanes contain translated hBVR. The molecular mass of the translated protein was approximated to be 69 kDa. This value was obtained using high molecular weight native markers. The third lane is that of the control, which consisted of rabbit reticulocyte lysate with all components present in the translation system minus hBVR mRNA.

differentiate between  $^{35}\text{S}$ -labeled protein and  $^{32}\text{P}$ -labeled DNA, the processed gel was exposed to two films separated by an opaque piece of paper, with an enhancing screen against the second film. This was to ensure that the film next to the gel was exposed to both  $^{35}\text{S}$  and  $^{32}\text{P}$ , while the film next to the screen was exposed only to higher energy  $^{32}\text{P}$  radiation. As shown in Fig. 5A, the translated hBVR did not bind to a 56-mer DNA fragment having one AP-1 site, while it did bind to the 100-mer

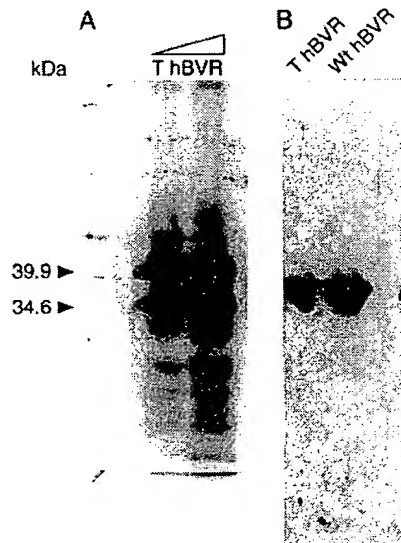
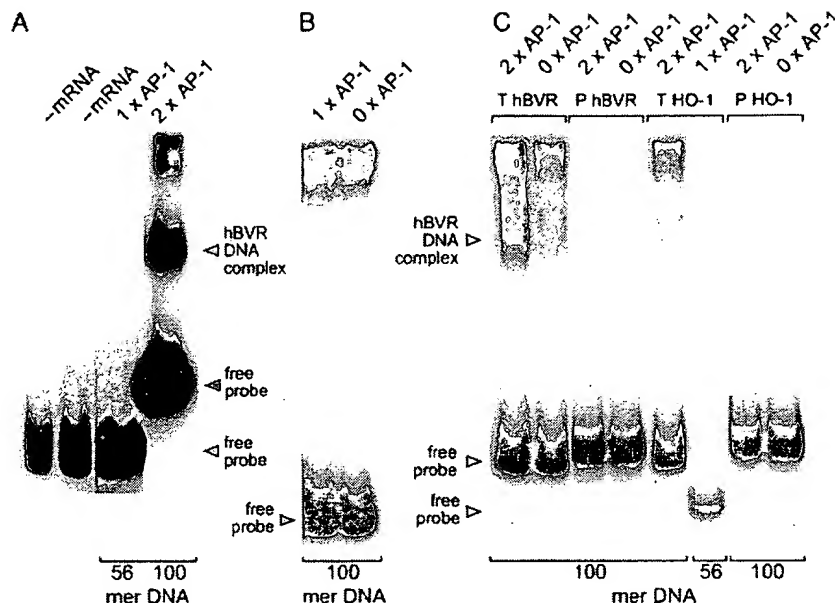


FIG. 4. Identification of the *in vitro* translated proteins as hBVR by Western blot analysis. A, SDS-polyacrylamide gel electrophoresis of *in vitro* translated BVR with two different amounts of lysate loaded. A 12% SDS gel was used for this experiment. The loading was not intended to be quantitative. Standard molecular mass protein markers indicated the apparent molecular mass of the translated protein bands being 39.9 and 34.6 kDa. B, Western blot analysis of *in vitro* translated hBVR. The first lane contained the translated hBVR; the second lane contained the wild type *E. coli* expressed purified hBVR. The primary antibody was rabbit anti-human kidney BVR. The difference in size of the images shown in A and B is due to the differential treatment of gels that were required for visualization of translated protein. T, *in vitro* translated hBVR; Wt, wild type hBVR.

DNA fragment having two AP-1 sites. For these experiments, the control contained labeled DNA with rabbit lysate minus hBVR mRNA. As noted in the figure, binding complexes were

**FIG. 5. hBVR DNA binding assay.** The binding assay was carried out using *in vitro* translated hBVR or HO-1 with modifications denoted for each lane. In A, the first two lanes from the left are controls containing the rabbit lysate but without hBVR mRNA. hBVR binding to the 56-mer DNA with one AP-1 site and binding to the 100-mer DNA fragment with two AP-1 sites are shown in the third and fourth lanes, respectively. The 56-mer DNA used in this experiment has been shown to bind with c-Jun/c-Fos heterodimer (26). The sequence of the 100-mer-long DNA fragment is that of the mouse HO-1 promoter region (39). B, analysis of hBVR binding to the 100-mer DNA fragment with one or zero AP-1 sites. C shows translated HO-1 (THO-1) to the 56- and 100-mer DNA fragments with one or two AP-1 sites, respectively. Also, binding of purified HO-1 to 100-mer DNA with two or zero AP-1 sites are shown. For comparison, binding of BVR to 100-mer DNA with two or zero AP-1 sites are shown.



not detectable in the control lanes. Also, binding of *E. coli* expressed hBVR protein, which is in monomeric form, to the 100-mer DNA fragment with two AP-1 sites was not detected.

Subsequently, the specificity of DNA binding and the number of AP-1 sites required for binding were examined. For this, hBVR-AP-1 binding was compared between three 100-bp DNA fragments with two, one, or zero AP-1 sites. As shown in Fig. 5B, hBVR binding requires two copies of the AP-1 binding sequence, because the interactions of hBVR with 100-bp fragments containing one or zero AP-1 sites were comparable, and the subdued signal appeared to reflect AP-1-unrelated DNA-protein interaction. To further examine the specificity of hBVR DNA binding, binding of *in vitro* translated HO-1 to the same AP-1-containing 56-mer and 100-mer DNA fragments was examined (Fig. 5C). The larger DNA had two AP-1 sites. Also, DNA binding was examined using *E. coli* expressed hHO-1 protein. As noted in Fig. 5C, neither the *in vitro* translated HO-1 nor the purified protein exhibited binding to the DNA fragments. The specificity of binding was assured by the addition of control unlabeled 100-mer DNA to all DNA binding experiments that used the 100-mer test DNA fragment. The control for the 56-mer test DNA was a 56-mer control unlabeled DNA fragment.

**In Vitro Translation of hBVR Leucine Zipper Mutants and Their Binding to DNA**—To establish the role of the leucine zipper motif of hBVR in DNA-protein interaction, site-directed mutagenesis studies were carried out. Mutations were directed to Lys<sup>143</sup>, Leu<sup>150</sup>, and Leu<sup>157</sup> that were changed to alanine, thereby generating K143A, L150A, and L157A, respectively. This was a particularly relevant investigation, because, as noted above, the model of the secondary structure of hBVR (Fig. 2) predicts a  $\beta$ -sheet structure for hBVR between Lys<sup>143</sup> and Leu<sup>157</sup>, while the structure common to most leucine zipper DNA-binding proteins is often entirely  $\alpha$ -helical. Studies with Jun and Fos oncoproteins suggest that single mutations in the motif are sufficient to abolish specific DNA binding (35). It has also been shown that a single amino acid change in Fos abolishes the DNA binding capabilities of the Fos-Jun dimer complex.

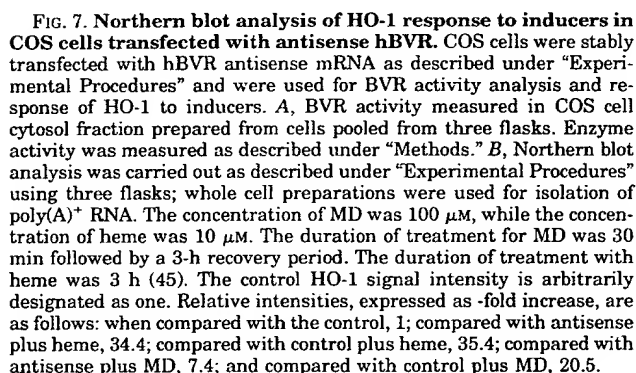
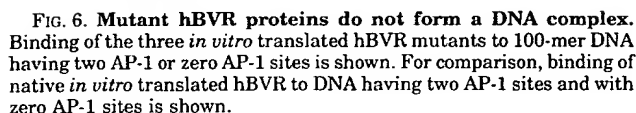
For this set of experiments, the [<sup>35</sup>S]methionine-labeled mutant BVR proteins were generated by *in vitro* translation and assayed on a 12% native gel for detection of the ~69-kDa protein band and analysis of DNA for complex formation. The

100-mer DNA fragment with two AP-1 sites or without an AP-1 site was used. On the native gel, the high molecular weight band was not detected with the mutated proteins. Also, as shown in Fig. 6, a single mutation in any of the three positions prevented protein-DNA complex formation. As noted, binding of the three mutant proteins with the DNA fragment having two or zero AP-1 sites was essentially comparable and was similar to that of the native hBVR binding to the 100-bp fragment with no AP-1 site. As before, the control, *in vitro* translated hBVR shows clear binding with DNA having two AP-1 sites.

The three-dimensional conformation of hBVR leucine zipper domain, predicted by the RasMol molecular graphic program (36), suggested that substitution of Leu<sup>143</sup>, Leu<sup>150</sup>, or Leu<sup>157</sup> by alanine in the leucine zipper motif apparently does not cause conformational changes in the motif and hence, most likely, does not account for the attenuated DNA binding.

**HO-1 Response to Menadione and Heme in COS Cells Transfected with Antisense hBVR**—To examine whether DNA binding of hBVR has any bearing on gene expression, induction of HO-1 in COS cells stably transfected with antisense hBVR was examined. HO-1 is transcriptionally regulated by a vast array of stimuli that trigger activation of different regulatory factors. MD and heme are both inducers of HO-1 gene expression but involve distinctly different signaling cascades-activating factors. To determine whether the antisense mRNA affected BVR activity, activity in the transfected cells was measured. As shown in Fig. 7A, a 66% decrease in activity was detected. This cell line was then used to examine the response of HO-1 to known inducers, heme and MD, by Northern blot analysis. As noted in Fig. 7B, the response of cells carrying antisense hBVR to heme did not differ from that of the control cells, and an increase of ~35-fold in HO-1 mRNA was detected in both sets of cells. In contrast, MD, which is a generator of oxygen radicals, produced a less than remarkable increase in HO-1 mRNA levels in the transfected cells. The control cells, on the other hand, displayed a robust response to MD. The magnitude of increase in HO-1 mRNA in the control and transfected cells was 20-fold versus 7-fold, respectively. HO-1 mRNA in COS cells with an absence of inducers was marginally detectable.





When a leucine zipper motif in the primary sequence of hBVR was detected, we considered that in BVR the motif could either be involved in dimerization, DNA binding, or some other functions related to its kinase activity. Of course, the possibility that the motif is of no apparent biological significance was not ruled out. A unifying feature of sequence-specific DNA-binding proteins is dimerization. Presently, evidence is provided that indicates formation of a homodimer by hBVR that binds to DNA and involves the leucine repeat region; the DNA binding sites are identified as two AP-1 recognition sequences. The finding that the single form of the nascent protein (Fig. 3) dissociates in two species (Fig. 4) under denaturing conditions and identification of the proteins based on their immunoreactivity as BVR (Fig. 4B) are indicative of a BVR homodimer formation. Moreover, the reductase contains the characteristic putative dimerization interface made of L<sub>1</sub>, L<sub>2</sub>, K<sub>3</sub>, L<sub>4</sub>, and L<sub>5</sub>, which is found in several proteins that bind nucleic acids (Fig. 1). The finding that site-directed mutation of these residues blocks the ability of hBVR to form a complex with 100-mer

Although hBVR has similarities in structure to a number of DNA-binding proteins with a leucine zipper motif, it also has divergent features. Moreover, based on the predicted secondary structure of hBVR, the sequence of amino acids between Leu<sup>129</sup> and Lys<sup>143</sup> forms an  $\alpha$ -helical structure, while the sequence between Lys<sup>143</sup> and Leu<sup>157</sup> is mainly  $\beta$ -sheet. Notably, the predicted secondary structure for many leucine zipper DNA-binding proteins is two  $\alpha$ -helices separated by a  $\beta$ -turn. Also, GCN4, a leucine zipper type DNA-binding protein falls short of such a helix-turn-helix motif (30). The DNA contact region in many of the leucine zipper proteins is the sequence immediately NH<sub>2</sub>-terminal to the leucine zipper with a notable degree of basicity that starts seven residues N-terminal to L<sub>1</sub>. In BVR, however, the content of basic amino acids in this region is low in comparison with that of other DNA-binding proteins, and unlike those proteins that have two clusters of basic residues linked by a spacer sequence with an invariant alanine spacer, only one basic cluster is present in BVR (Fig. 1). A second N-terminal basic domain is also absent from c-Myc, which is a helix-loop-helix DNA-binding protein. It has, however, a basic domain near the C terminus of the protein. The reductase has a basic domain near the carboxyl terminus of the protein: KKRILH (residues 275–280), which plausibly could also interact with DNA. In addition, the second basic domain is also absent in the leucine zipper protein human Shaker K<sup>+</sup> channel 3  $\beta$ -subunit (Fig. 1), which interestingly is also an oxidoreductase (38). In Shaker, which is a member of the aldo-ketoreductase superfamily, the leucine zipper motif is involved in interaction of K<sup>+</sup> channel subunits and to our knowledge has not been reported to bind to DNA.

On the basis of the denoted observations, it is reasonable to suspect that BVR may have a function of sorts in the AP-1 pathway of cell signaling. MD has long been used as an oxidative stress model. It stimulates the rate of NADPH oxidation,  $\text{H}_2\text{O}_2$  production, and redox cycling that results in formation of superoxide anions (44). The previous findings, that the reductase is activated by the oxidant,  $\text{H}_2\text{O}_2$ , and is a serine/threonine kinase (1), lend support to this idea. Noteworthy is that  $\text{H}_2\text{O}_2$  is an activator of HO-1 gene expression (45, 46). The suggestion

that hBVR-DNA binding is linked to the activation of the HO-1 gene is also consistent with previous observations that, in HeLa cells in response to cGMP and in intact rats in response to lipopolysaccharides or to the free radical generating compound, bromobenzene, reductase translocates from the cytosol to the nucleus (2). All mentioned stimuli are inducers of HO-1 gene expression.

**Acknowledgment**—We are grateful to Suzanne Bono for preparation of the manuscript.

## REFERENCES

- Salim, M., Brown-Kipphut, B. A., and Maines, M. D. (2001) *J. Biol. Chem.* **276**, 10929–10934
- Maines, M. D., Ewing, J. F., Huang, T. J., and Panahian, N. (2001) *J. Pharmacol. Exp. Ther.* **296**, 1091–1097
- Fakhrai, H., and Maines, M. D. (1992) *J. Biol. Chem.* **267**, 4023–4029
- Maines, M. D., Plevoda, B. V., Huang, T. J., and McCoubrey, W. K., Jr. (1996) *Eur. J. Biochem.* **235**, 372–381
- Kutty, R. K., and Maines, M. D. (1981) *J. Biol. Chem.* **256**, 3956–3962
- Schluchter, W. M., and Glazer, A. N. (1997) *J. Biol. Chem.* **272**, 13562–13569
- Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) *Science* **240**, 1759–1764
- Vinson, C. R., Sigler, P. B., and McKnight, S. L. (1989) *Science* **246**, 911–916
- O'Shea, E., Rutkowski, R., and Kim, P. S. (1989) *Science* **243**, 538–542
- Struhl, K. (1989) *Trends Biochem. Sci.* **14**, 137–140
- Ramirez-Carrozzi, V. R., and Kerppola, T. K. (2001) *J. Biol. Chem.* **276**, 21797–21808
- Kobe, B., and Dieneshofer, J. (1993) *Nature* **366**, 751–756
- Kobe, B., and Dieneshofer, J. (1994) *Trends Biochem. Sci.* **19**, 415–421
- Maines, M. D., Trakshel, G. M., and Kutty, R. K. (1986) *J. Biol. Chem.* **261**, 411–419
- Lee, P. J., Camhi, S. L., Chin, B. Y., Alam, J., and Choi, A. M. K. (2000) *Am. J. Physiol.* **279**, L175–L182
- He, C. H., Gong, P., Hu, B., Stewart, D., Choi, M. E., Choi, A. M. K., and Alam, J. (2001) *J. Biol. Chem.* **276**, 20858–20865
- Angel, P., and Karin, M. (1991) *Biochim. Biophys. Acta* **1072**, 129–157
- Han, Z., Boyle, D. L., Chang, L., Bennett, B., Karin, M., Yang, L., Manning, A. M., and Firestein, G. S. (2001) *J. Clin. Invest.* **108**, 73–81
- Devary, Y., Gottlieb, R. A., Lau, L. F., and Karin, M. (1991) *Mol. Cell. Biol.* **11**, 2804–2811
- Kikuchi, A., Park, S. Y., Miyatake, H., Sun, D., Sato, M., Yoshida, T., and Shiro, Y. (2001) *Nat. Struct. Biol.* **8**, 221–225
- Jackson, R., and Hunt, T. (1983) *Methods Enzymol.* **96**, 50–74
- Yoshida, T., Biro, P., Cohen, T., Müller, R. M., and Shibahara, S. (1988) *Eur. J. Biochem.* **71**, 457–464
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Cohen, S. N., Chang, A. C. Y., and Hsu, L. (1972) *Proc. Natl. Acad. Sci. U. S. A.* **69**, 2110–2114
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) *Science* **239**, 487–491
- Halazonetis, T. D., Georgopoulos, K., Greenberg, M. E., and Leder, P. (1988) *Cell* **55**, 917–924
- Maines, M. D., and Trakshel, G. M. (1993) *Arch. Biochem. Biophys.* **300**, 320–326
- Shibahara, S., Muller, R. M., and Taguchi, H. (1987) *J. Biol. Chem.* **262**, 12889–12892
- Ewing, J. F., and Maines, M. D. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 5364–5368
- Hope, I. A., and Struhl, K. (1985) *Cell* **43**, 177–188
- Alam, J., Cai, J., and Smith, A. (1994) *J. Biol. Chem.* **269**, 1001–1009
- Busch, S. J., and Sassone-Corsi, P. (1990) *Trends Genet.* **6**, 36–40
- Johnson, P. F., and McKnight, S. L. (1989) *Annu. Rev. Biochem.* **58**, 799–839
- Huang, T. J., Trakshel, G. M., and Maines, M. D. (1989) *J. Biol. Chem.* **264**, 7844–7849
- Ransone, L. J., Visvader, J., Sassone-Corsi, P., and Verma, I. M. (1989) *Genes Dev.* **3**, 770–781
- Ahmad, Z., and Sherman, F. (2001) *J. Biol. Chem.* **276**, 18450–18456
- Hope, I. A., and Struhl, K. (1986) *Cell* **46**, 885–894
- McCormack, T., and McCormack, K. (1994) *Cell* **79**, 1133–1135
- Alam, J., and Zhining, D. (1992) *J. Biol. Chem.* **267**, 21894–21900
- Devary, Y., Gottlieb, R. A., Smeal, T., and Karin, M. (1992) *Cell* **71**, 1081–1091
- Minden, A., Lin, A., Claret, F. X., Abo, A., and Karin, M. (1995) *Cell* **81**, 1147–1157
- Maines, M. D. (1992) *Heme Oxygenase-Clinical Applications and Functions*, CRC Press, Inc., Boca Raton, FL
- Lavrovsky, Y., Schwartzman, M. L., Levere, R. D., Kappas, A., and Abraham, N. G. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5987–5991
- Gillette, J. R., Brodie, B. B., and LaDu, B. N. (1957) *J. Pharmacol. Exp. Ther.* **119**, 532–540
- Keyse, S. M., Applegate, L. A., Tromvoukis, Y., and Tyrrell, R. M. (1990) *Mol. Cell. Biol.* **10**, 4967–4969
- Keyse, S. M., and Tyrrell, R. M. (1990) *Carcinogenesis* **5**, 787–791
- Van Straaten, F., Muller, R., Curran, T., Van Beveren, C., and Verma, I. M. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 3183–3187
- Gazin, C., Dupont de Dinechin, S., Hampe, A., Masson, J. M., Martin, P., Stehelin, D., and Galibert, F. (1984) *EMBO J.* **3**, 383–387
- Hinnebusch, A. G. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81**, 6442–6446
- Bohmann, D., Bos, T. J., Admon, A., Nishimura, T., Vogt, P. K., and Tjian, R. (1987) *Science* **238**, 1386–1392
- Hoeffler, J. P., Meyer, T. E., Yun, Y., Jameson, J. L., and Habener, J. F. (1988) *Science* **242**, 1430–1433
- Moye-Rowley, W. S., Harshman, K. D., and Parker, C. S. (1989) *Genes Dev.* **3**, 283–292
- Grupe, A., Schroter, K. H., Ruppertsberg, J. P., Stocker, M., Drewes, T., Beckh, S., and Pongs, O. (1990) *EMBO J.* **9**, 1749–1756



## Homodimerization via a Leucine Zipper Motif Is Required for Enzymatic Activity of Quiescent Cell Proline Dipeptidase\*

Received for publication, June 21, 2000

Published, JBC Papers in Press, June 23, 2000, DOI 10.1074/jbc.M005445200

Murali Chiravuri, Henry Lee, Suzanne L. Mathieu, and Brigitte T. Huber†

From the Department of Pathology, Program in Immunology, Tufts University School of Medicine, Boston, Massachusetts 02111

Quiescent cell proline dipeptidase (QPP) is an intracellular serine protease that is also secreted upon cellular activation. This enzyme cleaves N-terminal Xaa-Pro dipeptides from proteins, an unusual substrate specificity shared with dipeptidyl peptidase IV (CD26/DPPIV). QPP is a 58-kDa protein that elutes as a 120–130-kDa species from gel filtration, indicating that it forms a homodimer. We analyzed this dimerization with *in vivo* co-immunoprecipitation assays. The amino acid sequence of QPP revealed a putative leucine zipper motif, and mutational analyses indicated that this leucine zipper is required for homodimerization. The leucine zipper mutants showed a complete lack of enzymatic activity, suggesting that homodimerization is important for QPP function. On the other hand, an enzyme active site mutant retained its ability to homodimerize. These data are the first to demonstrate a role for a leucine zipper motif in a proteolytic enzyme and suggest that leucine zipper motifs play a role in mediating dimerization of a diverse array of proteins.

Quiescent cell proline dipeptidase (QPP)<sup>1</sup> is a 58-kDa protein that was recently isolated and cloned from human T cells (1). Highly specific inhibitors of post-proline cleaving aminodipeptidases cause cell death in quiescent lymphocytes, and the search for the target of these inhibitors led to the cloning of QPP and its subsequent nomenclature (2). QPP is a serine protease that cleaves dipeptides off the N terminus of proteins when the penultimate amino acid is a proline or an alanine. Although the substrates of QPP have yet to be elucidated, there are a striking number of cytokines, chemokines, and other signal molecules with highly conserved Xaa-Pro and Xaa-Ala motifs on the N terminus, rendering them potential substrates for QPP. Dipeptidyl peptidase IV (CD26/DPPIV), which shares substrate specificity with QPP, cleaves N-terminal Xaa-Pro motifs from chemokines such as macrophage-derived chemokines, regulated on activation normal T cell expressed and secreted, and stromal-derived factor 1 (3–5). This cleavage results in the functional inactivation of the three signal mole-

cules, indicating that this may be an important site of regulation of signal molecules *in vivo*.

Despite the large number of signal molecules with a conserved N-terminal Xaa-Pro motif, there are relatively few exopeptidases with the ability to cleave peptide bonds containing proline (6). These include the aminodipeptidases QPP and CD26/DPPIV and prolylcarboxypeptidase (PCP, angiotensinase), a post-proline cleaving carboxypeptidase that cleaves amino acids off the C terminus of proteins (7). The post-proline cleaving enzymes are likely to emerge as an important protease family. As indicated above, CD26/DPPIV has been shown to modulate the function of several chemokines (3–5), whereas PCP is a candidate gene for mediating essential hypertension (8). It is interesting to note that even though QPP and CD26/DPPIV share substrate specificity, they do not have homologous amino acid sequences. QPP shares a significant degree of homology with PCP (7) at the amino acid level (41% sequence identity) but not at the nucleotide level. QPP, CD26/DPPIV, and PCP share common structural features that may be reflective of their convergent evolution to form efficient post-proline cleaving enzymes: 1) they have the same ordering of the catalytic triad: Ser, Asp, His (1, 6); 2) they are glycoproteins, and glycosylation is essential for the enzymatic activity of at least two of them, CD26/DPPIV (9) and QPP<sup>2</sup>; and 3) CD26/DPPIV and PCP form homodimers, and we show here that QPP also oligomerizes. CD26/DPPIV forms homodimers through disulfide links (10, 11), whereas PCP forms homodimers through a poorly understood mechanism, believed to be mediated through serine repeats (12).

Leucine zipper motifs are protein-protein dimerization motifs consisting of heptad repeats of leucine residues that form a coiled-coil structure (13, 14). These motifs have been well described in the context of transcription factors such as c-Fos and c-Jun where they mediate homo- and heterodimerization critical for the DNA binding properties of these transcription factors (15). However, an increasing body of literature indicates that leucine zipper motifs mediate dimerization in a variety of other proteins. These include enzymes such as mixed lineage kinase-3 and tyrosine hydroxylase (16, 17), where the dimerization mediated by leucine zippers can be important for the activity of these enzymes. Mutagenesis has been used to analyze potential leucine zipper-mediated dimerization of a number of proteins (16–20).

We show here that active QPP elutes from gel filtration as a 120–130-kDa species, even though its estimated molecular mass from SDS-PAGE is 58 kDa (1). Primary sequence analysis of QPP revealed a putative leucine zipper motif upstream of the catalytic region (see Fig. 1). In this paper, we investigated the role of this putative leucine zipper in QPP homodimerization.

\* This work was supported by National Institutes of Health Research Grants AI36696 and AI43469 (to B. H.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Dept. of Pathology, Tufts University School of Medicine, 136 Harrison Ave., Boston, MA 02111. Tel.: 617-636-3989; Fax: 617-636-0449; E-mail: bhuber@opal.tufts.edu.

<sup>1</sup> The abbreviations used are: QPP, quiescent cell proline dipeptidase; DPPIV, dipeptidyl peptidase IV; PCP, prolylcarboxypeptidase; IP, immunoprecipitate; PAGE, polyacrylamide gel electrophoresis; HA, hemagglutinin; PBS, phosphate-buffered saline; AFC, amino-4-trifluoromethylcoumarin.

<sup>2</sup> M. Chiravuri, F. Agarraberes, S. L. Mathieu, H. Lee, and B. T. Huber, submitted for publication.

We used an *in vivo* co-immunoprecipitation scheme to analyze the dimerization properties of QPP. Independent point mutations in the leucine zipper region result in a loss of QPP homodimerization. The active site of QPP does not affect homodimerization, and a QPP active site mutant retains its ability to dimerize with wild type QPP, without influencing the enzymatic activity of wild type QPP. On the other hand, the leucine zipper mutants showed a loss of enzymatic activity. These results suggest that QPP homodimerizes through its putative leucine zipper motif and that this homodimerization is required for its enzymatic activity. These results may reflect the general requirement for structural features such as dimerization for post-proline cleaving enzymes. Furthermore, this is the first reported case of a leucine zipper motif mediating dimerization of a proteolytic enzyme.

#### MATERIALS AND METHODS

**Eukaryotic Expression Constructs**—The QPP constructs were cloned into the pCI-neo expression vector (Promega, Madison, WI), as described previously (1). The QPP-HA and QPP-Myc constructs were generated by polymerase chain reaction with the high fidelity DeepVent polymerase (New England Biolabs, Beverly, MA), using antisense primers incorporating either the Myc (EQLLISEEDL) or the HA (YPYDVP-DYA) epitope tags. To generate QPP mutants, the QuikChange site-directed mutagenesis kit (Stratagene, CA) was used. Briefly, primers containing the desired mutation together with flanking regions (>10 base pairs) were designed. Using QPP-HA in pCI-neo as a template, one round of polymerase chain reaction was performed with *Pfu* Turbo DNA polymerase. The products were subjected to digestion with *DpnI*, and the nicked vector DNA incorporating the mutation was transformed into XL1-Blue supercompetent cells.

**Transfections**—293T cells were grown in general medium (Dulbecco's modified Eagle's medium (Life Technologies, Inc.) with 10% fetal calf serum, 100 IU penicillin, 10 mg/ml streptomycin, L-glutamine, 2-mercaptoethanol, and sodium pyruvate). Transfections of 293T fibroblasts were performed using the calcium phosphate precipitation method. Two million cells/10-cm plate were plated 24 h prior to transfection. During transfection, 62  $\mu$ l of 2 molar  $\text{CaCl}_2$  was added to 438  $\mu$ l of double distilled  $\text{H}_2\text{O}$  containing the DNA (30  $\mu$ g) to be transfected. This was added to 500  $\mu$ l of 2 $\times$  HBS by bubbling. The mixture was immediately added to the cells. The medium was replaced after 8 h. Cells were harvested 48 h later. Stable 293T lines were generated by co-transfection of a QPP-HA construct in pCI-neo and pBABE-puro at a ratio of 15:1, respectively. These transfectants were expanded into general medium supplemented with 1.5  $\mu$ g/ml puromycin 48 h after transfection. Clones were selected and assayed for QPP expression.

**Gel Filtration Chromatography**—Stable QPP-HA expressing 293T cells were resuspended in 2.5 ml of lysis buffer (0.02 M phosphate buffer, pH 7.4, 4  $\mu$ g/ml aprotinin, 8  $\mu$ g/ml leupeptin, 8  $\mu$ g/ml antipain) and lysed by Dounce homogenization. The homogenate was centrifuged at 1000  $\times g$  for 10 min at 4  $^\circ\text{C}$ . The supernatant was then spun at 45,000  $\times g$  for 20 min and finally at 100,000  $\times g$  (S-110) for 1 h at 4  $^\circ\text{C}$ . The S-110 fraction was dialyzed for 16 h against 50 mM phosphate buffer + 150 mM NaCl at 4  $^\circ\text{C}$ . The S-110 fraction was loaded onto a Sephacryl S-200 column (Amersham Pharmacia Biotech) that was previously calibrated using a molecular mass marker kit (Sigma). The column was washed with 10 column volumes of column buffer (50 mM phosphate buffer + 150 mM NaCl). The column was run at 16 ml/h. 1.2-ml fractions were collected, assayed for Ala-Pro-AFC cleavage, and analyzed by Western blot analysis.

**Western Blot Analysis**—1–2  $\times 10^7$  cells were resuspended in lysis buffer (20 mM Hepes, 1.5 mM  $\text{MgCl}_2$ , 2 mM EDTA, 10 mM KCl, 0.1–1% Nonidet P-40, 5  $\mu$ g/ml antipain, and 5  $\mu$ g/ml leupeptin) for 30 min at 4  $^\circ\text{C}$ . The nuclei were spun out at 2000 rpm on a microcentrifuge for 10 min. Protein concentration was measured using the BCA protein estimation kit (Pierce). These samples were boiled for 5 min, subjected to SDS-PAGE, transferred onto polyvinylidene difluoride membranes, and blocked with 5% nonfat milk in PBS-T (PBS, 0.1% Tween-20y) for 1 h at room temperature. The primary anti-Myc (BD-Pharmingen, San Diego, CA) or anti-HA (BabCo, Richmond, CA) antibodies were incubated for 1 h at room temperature or overnight at 4  $^\circ\text{C}$  on an orbital shaker. The membrane was washed 3  $\times$  10 min with PBS-T. The secondary antibody, conjugated to horseradish peroxidase (Amersham Pharmacia Biotech) was incubated for 1 h at room temperature. The membrane was then washed 3  $\times$  15 min with PBS-T. The membrane was rinsed with

PBS, followed by addition of chemiluminescence substrate and autoradiography.

**Enzyme Assays**—1–2  $\times 10^7$  cells were resuspended in lysis buffer (20 mM Hepes, 1.5 mM  $\text{MgCl}_2$ , 2 mM EDTA, 10 mM KCl, 0.1% Nonidet P-40, 5  $\mu$ g/ml antipain, and 5  $\mu$ g/ml leupeptin) for 30 min at 4  $^\circ\text{C}$ . The nuclei were spun out at 2000 rpm on a microcentrifuge for 10 min. Protein concentration was measured, using the BCA protein estimation kit (Pierce). Lysates were added to a 96-well plate, followed by addition of the substrate solution (20  $\mu$ M Ala-Pro-AFC (Enzyme Systems Products, CA) in 50 mM Hepes). The samples were analyzed on a Fmax fluorescence plate reader (Molecular Devices) (Excitation, 390 nm; emission, 510 nm).

**Co-immunoprecipitation**—48 h after transfection, cells were washed in cold PBS and lysed in IP buffer (20 mM Hepes, 1.5 mM  $\text{MgCl}_2$ , KCl, EDTA, 1% Nonidet P-40, 5  $\mu$ g/ml leupeptin, and 5  $\mu$ g/ml phenylmethylsulfonyl fluoride) at 4  $^\circ\text{C}$  for 30 min. This was followed by a 300  $\times g$  centrifugation at 4  $^\circ\text{C}$ . The post-nuclear supernatant was transferred to a fresh Eppendorf tube, and the protein concentration was measured by BCA analysis. Samples were equalized for amount of protein and volume and then precleared by incubation with 50  $\mu$ l of protein G beads (Pierce) for 1 h at 4  $^\circ\text{C}$ . The post-nuclear supernatant was then treated with anti-HA antibody (6–8  $\mu$ g) overnight at 4  $^\circ\text{C}$  with shaking. 100  $\mu$ l of protein G beads were added for 1 h at 4  $^\circ\text{C}$ , and these beads were washed 4–5 times in lysis buffer. Finally, the beads were resuspended in 60  $\mu$ l of SDS loading buffer and boiled for 5 min. The samples were centrifuged, and the supernatants were run on SDS-PAGE and analyzed by Western blot.

#### RESULTS

**QPP Forms Dimers**—QPP is a 492-amino acid glycoprotein that is synthesized with a signal peptide (Fig. 1A). The active site serine is found in a consensus GXSYG sequence and together with Asp<sup>418</sup> and His<sup>442</sup> makes up the catalytic triad (1). These residues are conserved in the QPP homologue, PCP (1, 7). When initially purified from T cells, QPP was observed to elute as a 120-kDa species, whereas SDS-PAGE revealed a 58-kDa species, suggesting that QPP exists as a dimer (1).

To confirm these findings, QPP was analyzed by gel filtration chromatography (Fig. 2). 293T cells (2A5) stably expressing QPP were lysed and fractionated on a precalibrated gel filtration column (Fig. 2A), and QPP was analyzed by both Western blot analysis and enzyme activity assays. As can be seen in Fig. 2B, little QPP is found in those fractions corresponding to molecular masses of 50–60 kDa, while the majority of QPP is found in fractions corresponding to a molecular mass higher than 120 kDa. QPP requires glycosylation for its enzyme activity, and it migrates under SDS-PAGE as a 58-kDa species in its glycosylated form.<sup>2</sup> This glycosylated form was found predominantly in fractions 63–67, which would include proteins with molecular masses of 117 to approximately 144 kDa. Additional analysis was performed by measuring QPP activity in each of the fractions. Enzyme activity analysis shows that the majority of QPP activity is found in fractions 65–68, with the peak activity found in fractions 66 and 67, corresponding to a molecular mass of between 120 and 130 kDa (126 kDa) (Fig. 2C). These data show that the functionally active form of QPP exists as a dimer *in vitro*.

To investigate the dimerization properties of QPP *in vivo*, we utilized a co-immunoprecipitation scheme employing different epitope-tagged forms of QPP. Two QPP expression constructs with either a C-terminal Myc epitope tag or an HA epitope tag were transfected individually or in combination into 293T cells. Following transfection, lysates from these cells were immunoprecipitated with an anti-HA antibody, followed by SDS-PAGE and Western blot analysis using an anti-Myc antibody. As can be seen in Fig. 3A, following anti-HA immunoprecipitation, QPP-Myc was only precipitated when cotransfected with QPP-HA. Anti-HA Western blot analysis on the immunoprecipitates (Fig. 3B, lane 2) shows that the anti-HA antibody does not directly immunoprecipitate QPP-Myc and that the QPP-HA that is immunoprecipitated with the anti-HA antibody (Fig. 3B,

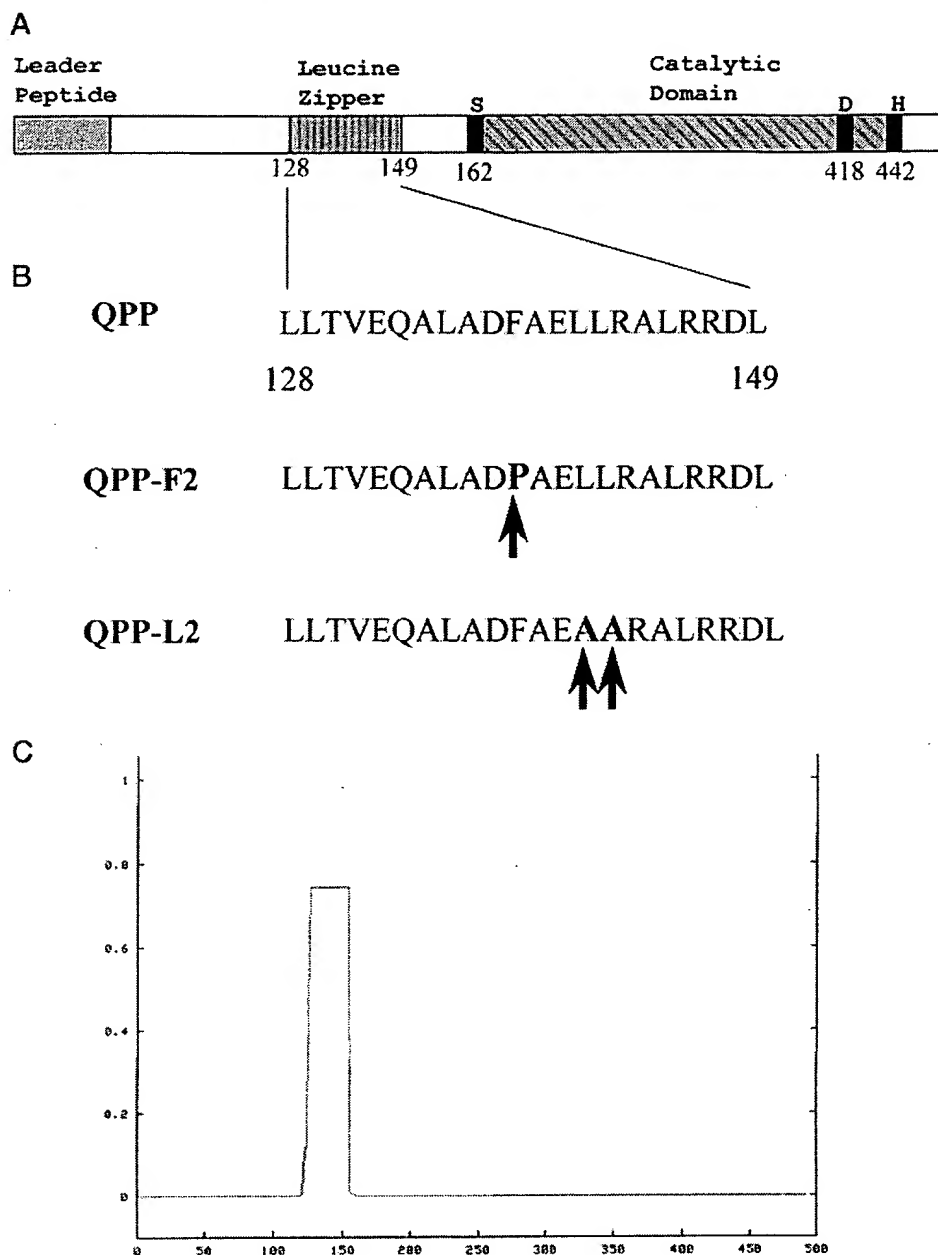


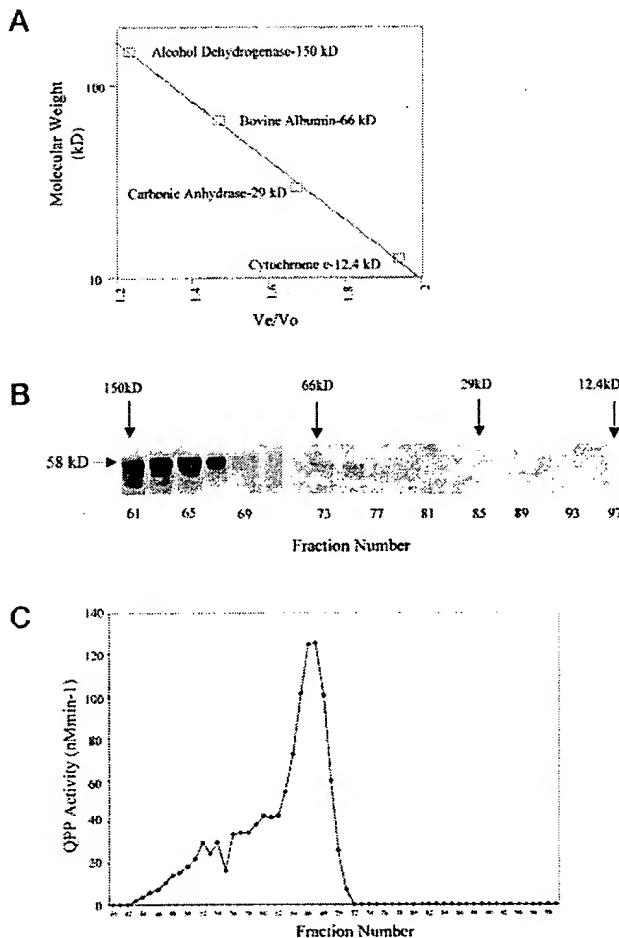
FIG. 1. Structure of QPP and QPP mutant constructs. A, schematic of QPP domain structure. B, sequence of the leucine zipper in wild type QPP (amino acids 128–149) and in the leucine zipper mutants QPP-F2 and QPP-L2. The mutations are shown in **bold** with arrows. C, the QPP sequence was analyzed by the COILS (23) program, using a 28-residue window.

lane 1) is not detected by the anti-Myc antibody (Fig. 3A, lane 1). This indicates that the anti-Myc and anti-HA antibodies are highly specific and do not cross react with the “opposing” epitope tag. Western blot analyses of the pre-IP lysates (Fig. 3, C and D) indicate that both constructs were expressed individually and in the cotransfected samples. These data show that QPP homodimerizes *in vivo* when overexpressed in 293T human fibroblast cells.

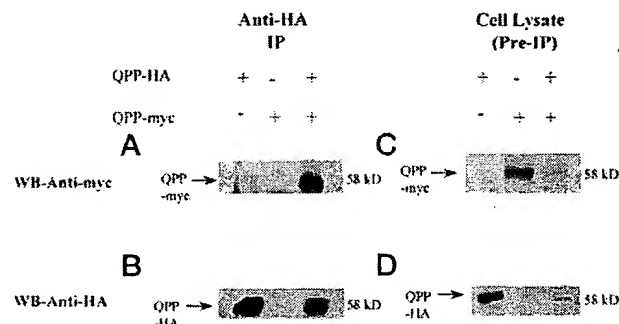
**Dimerization of QPP Is Mediated through a Leucine Zipper Motif**—Given that QPP forms oligomers, we analyzed the primary sequence of QPP for potential dimerization motifs. This analysis revealed a putative leucine zipper coiled-coil motif, consisting of a heptad repeat of leucine residues upstream of the catalytic domain (Fig. 1, B and C). Leucine zipper motifs serve as protein-protein interaction domains and mediate homo- and heterodimerization of a number of proteins (15–17,

19–22). To analyze the role of the putative leucine zipper in QPP, we designed two independent QPP leucine zipper mutants, QPP-F2 and QPP-L2. Both mutants were made using point mutations to minimize gross structural change in the mutant proteins. The first mutant, QPP-F2, was made by mutating Phe<sup>138</sup> to a proline. This mutation would be expected to disrupt the formation of the secondary  $\alpha$ -helical structure and thereby prevent formation of the quaternary coiled-coil structure. The second mutant, QPP-L2, involved the mutation of two leucine residues to two alanine residues (Leu<sup>141</sup> and Leu<sup>142</sup> to Ala). This more subtle mutation would not be expected to disrupt the secondary  $\alpha$ -helical structure of QPP, but the shorter alanine side chain would not be able to form the quaternary leucine zipper-mediated coiled-coil structure.

The same co-immunoprecipitation analyses as outlined above (Fig. 3) were carried out to test the ability of the leucine

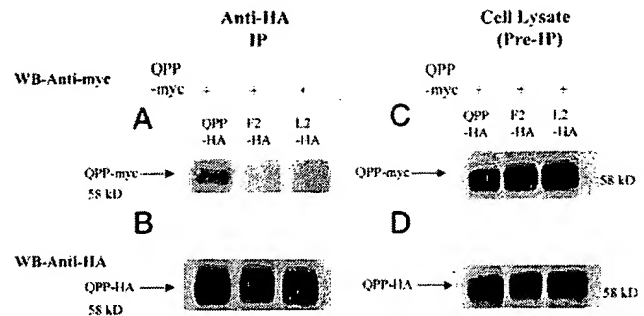


**FIG. 2. Gel filtration analysis of QPP.** QPP-HA expressing 293T cells were fractionated as described under "Materials and Methods." *A*, calibration of the Sephacryl S-200 gel filtration column with standard molecular mass markers.  $V_e/V_o$  denotes elution volume/void volume. *B*, Western blot analysis of the fractions using an anti-HA antibody. *C*, analysis of QPP enzyme activity of the various fractions.



**FIG. 3. QPP forms a homodimer.** *A*, cells were individually transfected with QPP-Myc, QPP-HA, or both constructs. Immunoprecipitation was performed with an anti-HA antibody, followed by SDS-PAGE and Western blot analysis (WB) using an anti-Myc antibody. *B*, Anti-HA Western blot analysis was performed on the anti-HA immunoprecipitates. *C*, Western blot analysis was performed on the pre-IP total lysates, using an anti-Myc antibody. *D*, Western blot analysis of the pre-IP total lysates, using an anti-HA antibody.

zipper mutants to form dimers. 293T cells were co-transfected with QPP-Myc and one of the following constructs: QPP-HA (wild type), QPP-F2-HA, or QPP-L2-HA. As seen in Fig. 4A, even as wild type QPP-HA homodimerized with QPP-Myc, both the QPP-F2 and the QPP-L2 leucine zipper mutants showed an



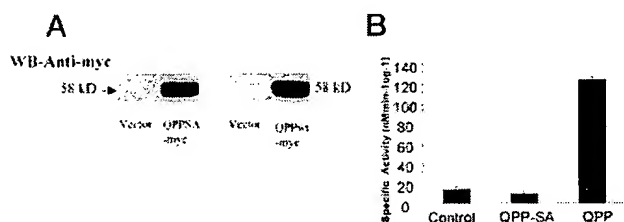
**FIG. 4. QPP homodimerization is mediated through a leucine zipper.** As in Fig. 2, the leucine zipper point mutants QPP-F2 and QPP-L2 were compared with wild type QPP for their ability to dimerize with QPP-Myc. Cells were co-transfected with QPP-Myc and QPP-HA, QPP-F2-HA, or QPP-L2-HA. *A*, anti-Myc Western blot analysis of anti-HA immunoprecipitates. *B*, anti-HA Western blot analysis of anti-HA immunoprecipitates. *C*, anti-Myc Western blot analysis of pre-IP cell lysates. *D*, anti-HA Western blot analysis of pre-IP cell lysates.

inability to homodimerize with the QPP-Myc construct. Anti-HA Western blot analysis of anti-HA immunoprecipitated lysates shows that the leucine zipper mutants immunoprecipitate at similar levels to wild type QPP (Fig. 4B), although the wild type QPP-Myc did not co-immunoprecipitate with the mutant constructs. Anti-Myc Western blot analysis of the pre-immunoprecipitate lysates indicated that all three samples shown in Fig. 4A had similar levels of QPP-Myc expression (Fig. 4C). Anti-HA Western blot analyses revealed that the leucine zipper mutants were expressed at similar levels as wild type QPP (Fig. 4D).

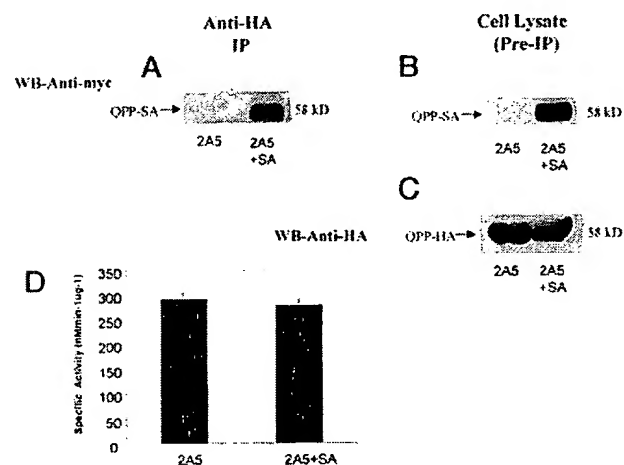
**A Functional Active Site Is Not Required for QPP Homodimerization**—QPP is a serine protease with an amino-dipeptidase activity that cleaves N-terminal dipeptides when the penultimate amino acid is a proline or an alanine. QPP overexpressed in 293T fibroblasts shows full functional activity in terms of its ability to cleave the reporter substrate Ala-Pro-AFC (Fig. 5B). The putative active site serine of QPP is found within a GXSG motif that is highly conserved between QPP and its homologues (1). A mutant was made that altered the active site serine into an alanine (GGAYG). This construct (QPP-SA) was expressed with a Myc epitope tag in 293T fibroblasts and was detected as a 58-kDa species by Western blot analysis using an anti-Myc antibody (Fig. 5A). However, this mutant (QPP-SA) showed no detectable enzymatic activity compared with the wild type QPP (Fig. 5B).

We used the active site mutant to determine whether 1) a functional active site is necessary for QPP dimerization and 2) in the event the active site mutant (QPP-SA) dimerized with wild type QPP, it affected QPP enzymatic activity. To answer these questions, we used a stable 293T line (2A5) expressing a QPP-HA construct that allowed us to analyze dimerization properties of the QPP active site mutant and to detect any effect of the QPP-SA construct on a fixed level of QPP-HA enzymatic activity of the 2A5 line.

The active site mutant QPP-SA with a Myc epitope tag was transfected into the 2A5 line (Fig. 6, A and B). Lysates from these cells were immunoprecipitated as before using an anti-HA antibody and subjected to SDS-PAGE and anti-Myc immunoblot analysis. As can be seen in Fig. 6A, the active site mutant retained its ability to dimerize with wild type QPP, indicating that a functional active site is not necessary for the dimerization to take place. 2A5 samples transfected with vector alone or the QPP-SA-Myc construct showed the same level of QPP-HA (Fig. 6C). We analyzed QPP enzyme activity in both of these samples, and, as seen in Fig. 6D, the active site mutant



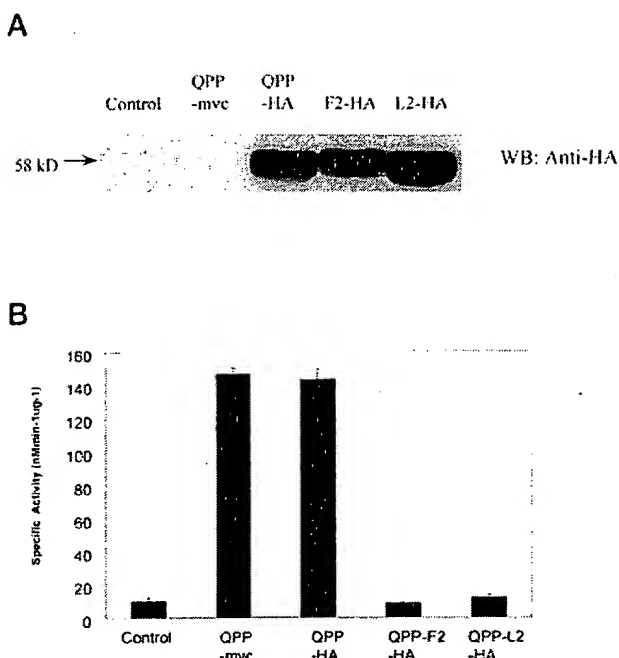
**FIG. 5. The active site mutant QPP-SA lacks enzymatic activity.** Wild type QPP and a serine-alanine mutant of QPP (QPP-SA) with a Myc epitope tag were expressed in 293T cells. *A*, anti-Myc Western blot analysis of QPP-SA and wild type QPP. *B*, analysis of enzymatic activity of the QPP-SA and wild type QPP constructs shown in *A*. QPP activity was measured with the fluorogenic substrate Ala-Pro-AFC. Control refers to lysates of vector transfected cells. WB, Western blot.



**FIG. 6. The active site of QPP does not play a role in QPP homodimerization.** The 2A5 line, which stably expresses QPP-HA, was transfected with vector alone or the QPP active site mutant QPP-SA (2A5 + SA). *A*, anti-Myc Western blot analysis of anti-HA immunoprecipitates. *B*, anti-Myc Western blot analysis of pre-IP samples. *C*, anti-HA Western blot analysis of pre-IP samples. *D*, analysis of QPP activity, measured by cleavage of Ala-Pro-AFC. WB, Western blot.

did not affect the enzymatic activity of the wild type QPP, even though the mutant dimerized with the wild type.

**The Leucine Zipper Is Critical for QPP Enzyme Activity**—We previously observed that QPP enzymatic activity was highly sensitive to structural alterations such as deglycosylation.<sup>2</sup> Given that homodimerization is a common feature of QPP, PCP, and CD26/DPPIV, we decided to investigate the effect of abolishing QPP dimerization on its activity. To determine the importance of the leucine zipper for QPP enzymatic activity, we measured enzymatic function of the QPP-F2 and QPP-L2 leucine zipper mutants of QPP that have mutations upstream of the catalytic triad. As seen in Fig. 7B, compared with vector transfected controls, the QPP-Myc and QPP-HA constructs showed full enzymatic activity. The QPP-F2 and QPP-L2 mutants, however, lacked enzymatic function. This was particularly interesting with the QPP-L2 mutant, which has a relatively subtle mutation of two leucines to two alanines. This construct, which also lacks the ability to homodimerize (Fig. 4), had no enzymatic activity over the vector transfected controls. To ensure equivalent expression levels in the samples tested, Western blot analysis was performed using an anti-HA antibody. As can be seen in Fig. 7A, the leucine zipper mutants QPP-F2 and QPP-L2 were expressed at similar levels as the wild type QPP-HA construct. QPP is a glycoprotein of 58 kDa that assumes a mass of 53 kDa when deglycosylated.<sup>2</sup> The fact that the QPP-F2 and QPP-L2 mutants migrate at 58 kDa (Fig. 7A) on SDS-PAGE analysis indicates that these mutants were



**FIG. 7. The leucine zipper mutants lose enzymatic activity.** *A*, 293T cells were individually transfected with vector, QPP-Myc, QPP-HA, QPP-F2-HA or QPP-L2-HA. Anti-HA Western blot (WB) analysis was performed to show expression of the HA-tagged constructs. *B*, analysis of QPP enzymatic activity of the different constructs.

correctly routed through the trans-Golgi network and underwent core and terminal glycosylation in the absence of homodimerization.

#### DISCUSSION

Leucine zipper motifs have been well described for transcription factors such as c-Jun where they mediate homo- and heterodimerization important for DNA binding (15). A number of reports, however, have shown that leucine zippers play a far more universal role as protein dimerization domains. For example, enzymes such as tyrosine hydroxylase (17), Mx1 GTPase (20), phenylalanine hydroxylase (21), human centrosomal kinase (Nek2) (22), and mixed lineage kinase (16) have been shown to dimerize through leucine zippers. QPP has a molecular mass of 58 kDa, as seen in SDS-PAGE, but elutes as a 120–130-kDa species in gel filtration (1), indicative of homodimer formation. Analysis of the QPP primary sequence revealed a leucine zipper heptad repeat structure that has a predicted ability to form a coiled-coil structure. This dimerization motif is upstream of the catalytic region of QPP. Interestingly, comparison of the deduced amino acid sequences of QPP cDNA derived from human and mouse shows that the leucine zipper motif is conserved, suggesting that this region is important for QPP activity.

In this report, *in vivo* co-immunoprecipitation assays were performed to analyze the dimerization properties of QPP. Two independent leucine zipper mutants lost their ability to homodimerize with wild type QPP. This suggests that the leucine zipper motif is important for homodimerization of QPP. The active site of QPP does not seem to play a role in homodimerization, because the active site mutant, QPP-SA, showed an ability to homodimerize with wild type QPP. Furthermore, the dimerization seems to be a structural rather than allosteric requirement, because dimerization with the QPP-SA mutant had no effect on the activity of wild type QPP. On the other hand, the leucine zipper mutants showed a complete loss of

enzymatic function. We, therefore, conclude that QPP homodimerization is mediated through a leucine zipper and that this homodimerization is required for QPP enzymatic activity, either to assume its correct structural conformation and/or to recognize and cleave its substrate. Disulfide bonds do not appear to play a role in QPP homodimerization as QPP migrates as a 58-kDa species under both reducing and nonreducing conditions (data not shown).

Two independent leucine zipper mutants yielded the same results in terms of dimerization and enzymatic activity; the QPP-F2 mutant is expected to lose secondary structure because of a kink introduced by the proline in the  $\alpha$ -helical structure, thus preventing the formation of a coiled-coil structure. On the other hand, in the QPP-L2 construct, the mutation of two leucine residues to two alanine residues would not be expected to alter the secondary  $\alpha$ -helical structure of QPP; thus, the results obtained with this mutant were particularly interesting. In both cases the mutations were introduced external to the catalytic domain and yet had profound effects on the enzymatic activity of QPP. The introduction of point mutations is not as drastic as a complete deletion of the entire leucine zipper, but we cannot entirely discount the possibility that such changes to the primary sequence cause changes in the folding pattern of QPP.

The post-proline cleaving exopeptidases, QPP, PCP, and CD26/DPPIV, have common structural features such as *N*-linked glycosylation and homodimerization that are important for their enzymatic activity (1, 6, 7, 9, 10, 12). Leucine zipper motifs play an important role in the dimerization of a wide array of proteins (15–17, 19–22). In this report we have described the first reported case of a functional requirement for a leucine zipper motif for the enzymatic activity of a proteolytic enzyme. These results will elucidate the mechanism of QPP enzymatic activity and post-proline cleaving exopeptidases in general. These results also confirm that leucine zippers mediate dimerization of diverse protein families.

**Acknowledgments**—We thank Dr. James Baleja for helpful discussions, Dr. Kurt Yardley, Nicole D'Avirro, and Sarada Tew for critical reading of the manuscript, and Lia Kim for excellent technical assistance.

## REFERENCES

- Underwood, R., Chiravuri, M., Yardley, K., Lee, H., Schmitz, T., and Huber, B. T. (1999) *J. Biol. Chem.* **274**, 34053–34058
- Chiravuri, M., Schmitz, T., Underwood, R., Yardley, K., and Huber, B. T. (1999) *J. Immunol.* **163**, 3092–3099
- Oravecz, T., Pall, M., Roderiquez, G., Gorrell, M. D., Ditto, M., Nguyen, N. Y., Boykins, R., Unsworth, E., and Norcross, M. A. (1997) *J. Exp. Med.* **186**, 1865–1872
- Shioda, T., Kato, H., Ohnishi, Y., Tashiro, K., Ikegawa, M., Nakayama, E. E., Hu, H., Kato, A., Sakai, Y., Liu, H., Honjo, T., Nomoto, A., Iwamoto, A., Morimoto, C., and Nagai, Y. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6331–6336
- Proost, P., Struyf, S., Schols, D., Opdenakker, G., Sozzani, S., Allavena, P., Mantovani, A., Augustyns, K., Bal, G., Haemers, A., Lambeir, A. M., Scharpe, S., Van Damme, J., and De Meester, I. (1999) *J. Biol. Chem.* **274**, 3988–3993
- Vanhoof, G., Goossens, F., De Meester, I., Hendriks, D., and Scharpe, S. (1995) *FASEB J.* **9**, 736–744
- Tan, F., Morris, P. W., Skidgel, R. A., and Erdos, E. G. (1993) *J. Biol. Chem.* **268**, 16631–16638
- Watson, B. Jr., Nowak, N. J., Myracle, A. D., Shows, T. B., and Warnock, D. G. (1997) *Genomics* **44**, 365–367
- Fan, H., Meng, W., Kilian, C., Grams, S., and Reutter, W. (1997) *Eur. J. Biochem.* **246**, 243–251
- Morimoto, C., and Schlossman, S. F. (1998) *Immunol. Rev.* **161**, 55–70
- von Bonin, A., Huhn, J., and Fleischer, B. (1998) *Immunol. Rev.* **161**, 43–53
- Skidgel, R. A., and Erdos, E. G. (1998) *Immunol. Rev.* **161**, 129–141
- Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) *Science* **240**, 1759–1764
- O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1989) *Science* **243**, 538–542
- Junius, F. K., O'Donoghue, S. I., Nilges, M., Weiss, A. S., and King, G. F. (1996) *J. Biol. Chem.* **271**, 13663–13667
- Leung, I. W., and Lassam, N. (1998) *J. Biol. Chem.* **273**, 32408–32415
- Vrana, K. E., Walker, S. J., Rucker, P., and Liu, X. (1994) *J. Neurochem.* **63**, 2014–2020
- Inoue, H., Takahashi, S., Fukui, K., and Miyake, Y. (1991) *J. Biol. Chem.* **266**, 11896–11900
- Simmerman, H. K., Kobayashi, Y. M., Autry, J. M., and Jones, L. R. (1996) *J. Biol. Chem.* **271**, 5941–5946
- Schumacher, B., and Staeheli, P. (1998) *J. Biol. Chem.* **273**, 28365–28370
- Huflon, S. E., Jennings, I. G., and Cotton, R. G. (1998) *Biochim. Biophys. Acta* **1382**, 295–304
- Fry, A. M., Arnaud, L., and Nigg, E. A. (1999) *J. Biol. Chem.* **274**, 16304–16310
- Lupas, A., Van Dyke, M., and Stock, J. (1991) *Science* **252**, 1162–1164

## A Heptad Motif of Leucine Residues Found in Membrane Proteins Can Drive Self-assembly of Artificial Transmembrane Segments\*

(Received for publication, October 2, 1998, and in revised form, January 15, 1999)

Rolf Gurezka, Rico Laage, Bettina Brosig, and Dieter Langosch†

From the Universität Heidelberg, Neurobiology Department, Im Neuenheimer Feld 364, 69120 Heidelberg, Germany

Specific interactions between  $\alpha$ -helical transmembrane segments are important for folding and/or oligomerization of membrane proteins. Previously, we have shown that most transmembrane helix-helix interfaces of a set of crystallized membrane proteins are structurally equivalent to soluble leucine zipper interaction domains. To establish a simplified model of these membrane-spanning leucine zippers, we studied the homophilic interactions of artificial transmembrane segments using different experimental approaches. Importantly, an oligoleucine, but not an oligoalanine, sequence efficiently self-assembled in membranes as well as in detergent solution. Self-assembly was maintained when a leucine zipper type of heptad motif consisting of leucine residues was grafted onto an alanine host sequence. Analysis of point mutants or of a random sequence confirmed that the heptad motif of leucines mediates self-recognition of our artificial transmembrane segments. Further, a data base search identified degenerate versions of this leucine motif within transmembrane segments of a variety of functionally different proteins. For several of these natural transmembrane segments, self-interaction was experimentally verified. These results support various lines of previously reported evidence where these transmembrane segments were implicated in the oligomeric assembly of the corresponding proteins.

In any type of cell, a multitude of integral membrane proteins is simultaneously synthesized and integrated into various membranes followed by association to homo- or heterooligomeric complexes. To ensure specific assembly, their subunits must present complementary recognition domains to each other. These domains may be located on the ectodomains and/or the transmembrane segments (TMSs).<sup>1</sup> Interactions between TMSs are currently intensely studied, since they usually form autonomous  $\alpha$ -helices and have been found to direct subunit assembly or support correct folding of many membrane proteins (1, 2). Biochemical and functional analyses, molecular modeling, and structural studies indicated that the self-assembly of transmembrane helices is driven by a close packing of

their characteristically shaped surfaces. These packing interactions may result in pairs of  $\alpha$ -helices with a right-handed twist as exemplified by glycophorin A (3, 4) and probably by synaptobrevin II (5). Other TMS interactions involve a leucine zipper type of side-chain packing as known from certain soluble proteins. Within soluble leucine zippers, the interacting residues form repeated heptad (*abcdefg*) motifs. Residues at *a*- and *d*-positions constitute the hydrophobic core of the interfaces; side-chains at the *e*- and *g*-positions are frequently charged, form salt bridges to each other, and make hydrophobic contacts to the core (6). Heptad motifs were also suggested to form the TMS interfaces of phospholamban (7, 8) and the M2 proton channel (9). Based on a quantitative evaluation of high resolution structures, we recently confirmed previous observations (10, 11) in demonstrating that TMSs primarily interact via a leucine zipper type of packing within bacteriorhodopsin, the photosynthetic reaction center, and cytochrome *c* oxidase. There, the heptads are repeated on average 2–3 times, and the motif *gaxxdexgaxxdexga* covers the central parts of the membrane-spanning interfaces. Salt bridges are absent due to the hydrophobic nature of most membrane-embedded residues (12).

To establish a simplified model of membrane-spanning leucine zipper domains, we designed artificial TMSs on the basis of leucine and alanine residues. We show that an oligoleucine sequence or a *gaxxdexgaxxdexga* motif of leucine residues elicits specific self-assembly in membranes and in detergent solution. Interestingly, variants of this motif are found within the TMSs of a diverse set of natural membrane proteins, where they appear to be important for oligomeric assembly.

### EXPERIMENTAL PROCEDURES

**Plasmid Constructs**—Construction of plasmids pToxRATM and pSNiRATM was described previously (5, 13). All other pToxR constructs were made by ligating synthetic oligonucleotide cassettes encoding the desired sequences into the plasmid pHKToxR(TM<sup>14</sup>)MalE (14) previously cut with *Nhe*I and *Bam*HI. For the nuclease A fusions, the oligonucleotide cassettes were ligated into plasmids pSNiR (5) or pSNiR2 previously cut with *Nhe*I and *Bam*HI. Details on the pSNiR and pSNiR2 plasmids will be described elsewhere. All constructs were verified by dideoxy sequencing.

**ToxR Activity Assays**—Transcription activation was determined upon expression of the pToxR constructs in the indicator strain FHK12 as described (15). 0.4 mM isopropyl 1-thio- $\beta$ -D-galactopyranoside was added to the cultures to enhance the dynamic range of the produced  $\beta$ -galactosidase signals (in Miller units (MU), means  $\pm$  S.D.) elicited by the different constructs in several independent experiments. This effect is thought to result from isopropyl 1-thio- $\beta$ -D-galactopyranoside-induced expression of an F'-plasmid-encoded truncated  $\beta$ -galactosidase, which competes with full-length enzyme in the formation of functional tetramers. The previously (15) described construct pToxR/GPA13 elicited  $1240 \pm 298$  MU under these conditions.

**Gel Filtration Chromatography**—pSNiR and pSNiR2 fusion proteins were expressed in BL21(DE3)pLysS cells (Novagen), solubilized in 25 mM HEPES, pH 7.9, 0.5 M NaCl, 2% CHAPS, 1 mM EDTA and quantitated as described (5). Volumes of 300  $\mu$ l at concentrations of 4 or 20  $\mu$ M fusion protein were separated on a Superdex 200HR 10/30 column

\* This work was supported by Deutsche Forschungsgemeinschaft (Grant La699/4-1 and Heisenberg Program) and the Fonds der Chemischen Industrie. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence and reprint requests should be addressed: Institut für Neurobiologie, Uni Heidelberg, Im Neuenheimer Feld 364, 69120 Heidelberg, Germany. Tel.: 06221-548696; Fax: 06221-544496; E-mail: Langosch@sun0.urz.uni-heidelberg.de.

<sup>1</sup> The abbreviations used are: TMS, transmembrane segment; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propane sulfonic acid; EpoR, erythropoietin receptor; MalE, maltose-binding protein; MU, Miller units.



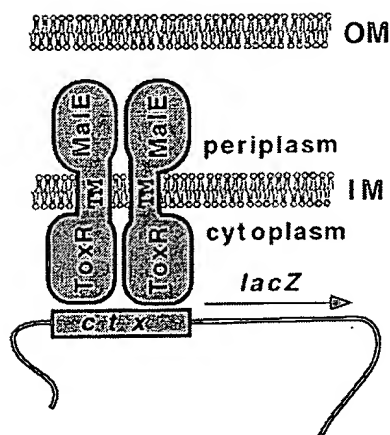


FIG. 1. Functional organization of ToxR chimeric proteins. The cytoplasmic domain ToxR is linked via a TMS of choice to the periplasmic MalE moiety. Upon dimerization, ToxR binds to the *ctx* promoter, thus initiating *lacZ* transcription in the indicator cells. TM, transmembrane segment; MalE, maltose-binding protein; OM, outer membrane; IM, inner membrane.

(Amersham Pharmacia Biotech FPLC system) using a flow rate of 0.5 ml/min and 25 mM HEPES, pH 7.9, 0.5 M NaCl, 1% CHAPS, 1 mM EDTA as running buffer. Fractions of 0.5 ml were collected and analyzed for fusion protein with a dot blot procedure (16) using the 9E10 monoclonal antibody directed against the *c-myc* marker epitope for detection. The elution profiles were constructed from the antigen content, and the apparent molecular weights were calculated with reference to standards given in the legend to Fig. 3.

**Data Base Searching**—The Swiss-Prot data base (release 35.0) was searched with the LLXXLLXXLLXXLL motif using the Findpatterns option of the HUSAR sequence analysis package made available by the German Cancer Research Center (Heidelberg). Up to three mismatches were allowed. To selectively retrieve TMSs, any amino acid except the charged residues lysine, arginine, glutamate, aspartate, or the helix-breaker proline was allowed for those positions not occupied by leucine.

**Miscellaneous Methods**—Western blotting was done as described with an antiserum recognizing the maltose-binding protein (MalE) moiety of the constructs, and the bands were quantitated densitometrically (13, 15). The ability of our constructs to complement the MalE deficiency of PD28 cells was tested by measuring the cell densities of transformed bacteria in minimal medium containing maltose at 640 nm after different growth periods (13). NaOH extraction was done as described (17) by vortexing whole bacteria with cold 0.1 M NaOH followed by centrifugation to separate soluble from membrane-bound proteins.

## RESULTS

**A Model of Membrane-spanning Leucine Zipper Domains**—Leucine is the most prevalent amino acid within the interface of leucine zippers (18), which is probably related to its ability to adopt multiple conformations (19). We therefore reasoned that the flexible leucine side chain may be particularly well suited to form a well packed membrane-spanning leucine zipper. The methyl side chain of alanine, in contrast, is expected to be too small for efficient interaction with other alanine residues. This prediction was tested by comparing the self-association of oligoleucine and oligoalanine sequences, which are known to form stable  $\alpha$ -helices (20, 21).

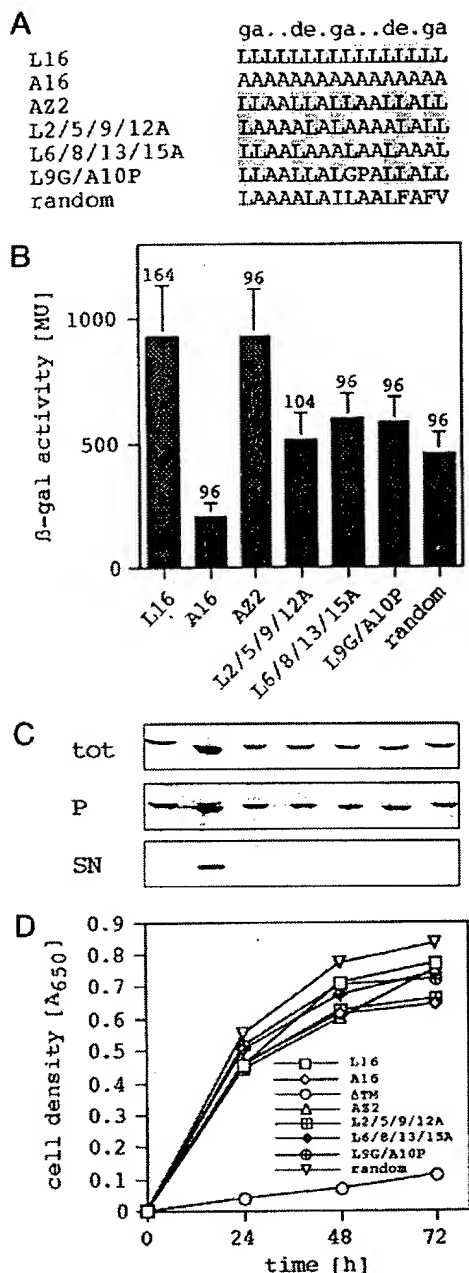
One of the experimental approaches we used is based on an engineered version of the ToxR transcription activator. This protein is anchored by a single TMS of choice within the inner membrane of expressing *Escherichia coli* cells, where it is thought to exist in a monomer/dimer equilibrium. The dimeric form binds to the cholera toxin promoter, thus activating expression of a downstream *lacZ* gene in a reporter strain (Fig. 1; Ref. 14).  $\beta$ -Galactosidase expression is therefore diagnostic of ToxR self-assembly in the membrane. We previously established this system as a sensitive tool to study TMS interactions using the structurally well characterized glycophorin A TMS

dimer for reference (13, 15).

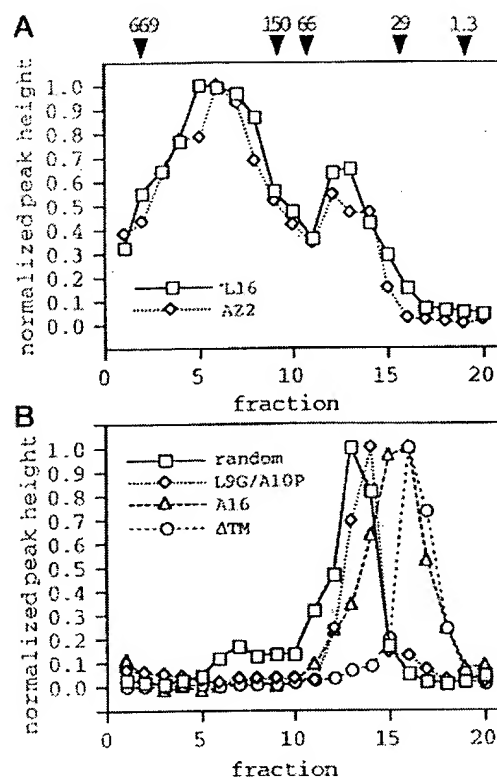
Here, we found that a sequence of 16 leucine residues (designated L16) elicited strong transcription activation ( $924 \pm 209$  MU; mean  $\pm$  S.D.). In contrast, a sequence of 16 alanine residues (designated A16), elicited only a weak signal ( $210 \pm 53$  MU) (Fig. 2, A and B). This suggests that the oligoleucine sequence self-assembles in the membrane, whereas the oligoalanine sequence stays largely monomeric. Thus, the latter can be used as host for a leucine zipper motif. Based on the *gaxxdexgaxxdexga* motif representing the central parts of most transmembrane helix-helix interfaces within crystallized membrane proteins (12), a simplified version of a membrane-spanning leucine zipper interaction domain was designed. In this model, the  $\alpha$ ,  $d$ ,  $e$ , and  $g$  positions are occupied by leucine and all others by alanine. The construct with this hybrid sequence (AZ2) self-interacted to a similar degree ( $929 \pm 186$  MU) as the parental L16 protein (Fig. 2, A and B). To demonstrate that the leucine residues contained within AZ2 constitute the helix-helix interface, we mutated some of them to alanine and assessed the consequences for self-interaction. None of the single mutations made (L2A, L5A, L9A) significantly reduced the signal (data not shown). However, when either four  $\alpha$  and  $d$  (L2A/L5A/L9A/L12A) or four  $g$  and  $e$  (L6A/L8A/L13A/L15A) positions were simultaneously mutated, the signal dropped by about 50% ( $516 \pm 106$  or  $596 \pm 102$  MU). Thus, the leucine residues are critical for the interaction and, hence, most likely make up the interface. Further,  $ad$ - and  $eg$ -positions seem to be of similar importance for helix-helix packing. Introducing a glycine-proline pair into the center of the AZ2 sequence (L9G/A10P) similarly affected the interaction ( $584 \pm 100$  MU), consistent with the known destabilization of  $\alpha$ -helices by glycine (22) and their kinking by proline (23) residues. We also replaced the leucines of AZ2 by three different random sequences consisting of the most abundant residues found within TMSs (leucine, isoleucine, valine, phenylalanine, alanine) (24) while maintaining total hydrophobicity and side-chain surface (25). Compared with AZ2, these random sequences also self-assembled much less efficiently, thus emphasizing the superior suitability of the leucine side chain for helix-helix packing (e.g. "random,"  $446 \pm 72$  MU; Fig. 2, A and B, and data not shown). The reductions in signal strength of the mutants compared with AZ2 are statistically highly significant (two-tailed Student's *t* test,  $p < 0.001$ ).

Comparing the concentrations of our ToxR constructs by Western blot analysis indicated that most of them were expressed at similar levels, whereas consistent overexpression was noted for the A16 construct (Fig. 2C). When we extracted the cells with NaOH to separate membrane proteins (pellet) from soluble proteins (alkali supernatant) (17), all constructs cosedimented quantitatively with the membranes as expected except A16, which could be partially alkali-extracted (Fig. 2C). Thus, a fraction of the A16 protein seems to remain in a soluble compartment, which is probably due to the comparably low hydrophobicity of the oligoalanine sequence. This fraction is thought not to interfere with the assay. To assess correct integration of the proteins into the inner membrane, we tested their ability to functionally complement the MalE deficiency of PD28 cells. Due to a MalE deletion, this *E. coli* strain is unable to grow in minimal medium with maltose as the only carbon source (26). In cells expressing correctly inserted ToxR membrane proteins with the ToxR moiety facing the cytoplasm and the MalE domain exposed to the periplasmic space (see Fig. 1), however, the MalE domain allows maltose uptake and thus cell growth (13, 14). Here, expression of all constructs including A16 complemented the MalE deficiency of PD28 cells to comparable degrees (Fig. 2D). In contrast, a control construct





**FIG. 2. Transcription activation, expression, and membrane incorporation of ToxR constructs with artificial TMSs.** A, TMS sequences aligned to the underlying heptad pattern. Leucine residues of the zipper variants are shaded for clarity. B, different levels of transcription activation elicited by the different constructs in FHK12 cells indicate sequence-specific TMS assembly in the membrane. The bars represent mean specific  $\beta$ -galactosidase activities calculated from numbers of data points given for each construct; error bars denote S.D. C, expression level and membrane association in FHK12 cells. tot, the total cell content of most ToxR proteins was similar as revealed by the staining intensities of the 65-kDa proteins upon Western blotting (densitometric quantitation of seven independent blots established that the average levels of the mutant TMSs ranged from 98 to 111% of the parental AZ2 protein), whereas ToxRA16 was overexpressed; P, the alkali-extracted membrane pellet quantitatively retained all constructs except ToxRA16; SN, the alkali supernatant contained part of ToxRA16 but none of the other proteins. The order of samples corresponds to that in B. D, functional complementation of MalE deficiency to assess correct membrane incorporation. All constructs except the control construct ToxRA $\Delta$ TM allowed for similar rates of PD28 cell growth, thus confirming their correct  $N_{in}$ - $C_{out}$  integration. The individual data points represent means from five independent experiments.



**FIG. 3. Oligomeric assembly in detergent solution.** Nuclease A fusion proteins were expressed, solubilized in CHAPS, and subjected to gel filtration chromatography at concentrations of 20  $\mu$ M. A, the L16 and AZ2 constructs assembled to  $\sim$ 300-kDa oligomeric complexes accompanied by 47-kDa minor peaks probably representing monomers. B, the major fractions of all other analyzed proteins migrated at apparent molecular weights consistent with monomers containing different amounts of bound detergent. Elution profiles are compared with the positions of marker proteins (vitamin B12, 1.35 kDa; carbonic anhydrase, 29 kDa; bovine serum albumin, 67 kDa; alcohol dehydrogenase, 150 kDa; thyroglobulin, 669 kDa). The chromatograms are normalized relative to their highest peaks.

where the TMS is deleted (ToxRA $\Delta$ TM) proved unable to support cell growth as expected from its presumed cytoplasmic localization. In sum, equivalent amounts of all ToxR proteins analyzed here for self-assembly appear to be correctly integrated into the inner bacterial membrane, and the obtained  $\beta$ -galactosidase activities can thus be directly compared.

To examine self-assembly of our artificial TMSs by an independent approach, their oligomeric states were directly compared in detergent solution (Fig. 3). The L16, A16, AZ2,  $\Delta$ TM, L9G/A10P, and "random" sequence segments were genetically fused to the C terminus of a fusion moiety based on *Staphylococcus aureus* nuclease A, a monomeric soluble protein. The fusion proteins were overexpressed in *E. coli*, solubilized with CHAPS, and subjected to gel filtration chromatography at concentrations of 4 or 20  $\mu$ M. When injected at 20  $\mu$ M, both L16 and AZ2 fusion proteins eluted as broad peaks with mean apparent molecular masses of  $\sim$ 300 kDa plus minor peaks at 47 kDa. At 4  $\mu$ M, the 300-kDa peaks were decreased in favor of the 47-kDa peaks, indicating equilibrium between both forms of the proteins (data not shown). Whereas the 300-kDa peaks clearly indicate assembly to multimers whose stoichiometry is currently not clear, the 47-kDa peaks most likely reflect monomers that may migrate at increased apparent molecular weights due to bound detergent (calculated molecular masses: L16, 21.2 kDa; AZ2, 20.9 kDa). In contrast to that, the  $\Delta$ TM, A16, L9G/A10P, and "random" constructs gave rise to major peaks at 17,

22, 31, and 41 kDa at both concentrations. These peaks are consistent with monomers (calculated masses: 19.5, 20.5, 20.9, and 20.9 kDa, respectively) whose migration may be influenced by different amounts of bound detergent depending on the presence and the hydrophobicity of the hydrophobic segments.

Taken together, two independent experimental approaches indicate that both the oligoleucine sequence and the model leucine zipper motif AZ2 self-assemble in a sequence-specific way in membranes as well as in detergent solution.

**Self-assembly of Leucine-rich Natural Transmembrane Segments**—Given the self-assembly of the AZ2 model, we assessed whether TMSs with similar leucine patterns exist in naturally occurring proteins. The Swiss-Prot data base was searched for hydrophobic sequence segments with the motif LLXXLLX-LLXXLLXLL allowing for up to three mismatches. This search yielded 38 predicted N-terminal signal sequences, 30 TMSs predicted within polytopic membrane proteins, and 15 predicted TMSs from bitopic membrane proteins when homologous proteins from different species were counted only once. Whereas the signal sequences and TMSs of polytopic proteins were not further investigated here, the TMS sequences corresponding to the bitopic proteins are shown in Table I. Self-interaction of a subset was examined with the ToxR system. The TMS eliciting the strongest signal was derived from the erythropoietin receptor followed by the TMSs of the Friend spleen focus-forming virus envelope protein, E-cadherin, and hemagglutinin of canine distemper virus. Other TMSs corresponding to papillomavirus E5 protein, mouse poliovirus receptor homolog, and chick asialoglycoprotein receptor gave rise to intermediate values suggesting lower levels of self-assembly (Fig. 4A and Table I). A Western blot run for control revealed roughly similar expression levels (Fig. 4B).

The data predict that these TMSs are important for oligomerization of the corresponding proteins. A survey of previously reported experimental evidence and our own experiments indicated this indeed to be the case for several of these proteins or related homologs as discussed below.

#### DISCUSSION

We demonstrate that an artificial TMS of leucine residues efficiently self-assembles in membranes and in detergent solution. A heptad motif of leucine residues suffices to elicit self-assembly, which therefore is thought to be driven by the type of side-chain packing known from leucine zipper interaction domains. The main implications of our results are 2-fold. (i) They establish a simplified model system of short membrane-spanning leucine zippers. (ii) They suggest that similar interaction domains may play a role in subunit-subunit recognition of certain natural membrane proteins.

#### Structural Aspects of Membrane-spanning Leucine Zippers

We assume that the L16 and AZ2 TMSs form  $\alpha$ -helical bundles upon self-assembly. Self-assembly is thought to involve self-complementary helix surfaces that associate which each other via a "knobs-into-holes" type of side-chain packing characteristic of leucine zippers (6). The highly flexible leucine side chain (19) may be particularly well suited for this type of packing interaction. Consistent with this concept, leucine-rich heptad motifs have previously been applied in the design of helix bundles forming transmembrane ion channels (27) or of a folded polytopic membrane protein (28). On the other hand, leucine helices have frequently been used as experimental models to study TMS interactions with lipid bilayers. For some of these studies (20, 29–31), both termini of the leucine helices were capped with lysine residues whose repulsive interaction may keep them in a monomeric state (31). In other cases (17, 32, 33), their self-assembly as implied by our data should be

TABLE I  
Heptad motifs of leucine residues in natural TMSs

Swiss-prot ID <sup>a</sup>	TMS sequence <sup>b</sup>	Activity <sup>c</sup>
	<i>gaxxdexgaxxdexga</i>	<i>MU</i>
CAD1_XENLA	<sup>705</sup> ILGGILALLLLLLLLL	854 ± 162
CAD3_HUMAN	<sup>659</sup> VLGAVLALLFLLLVLL	ND <sup>d</sup>
CADB_CHICK	<sup>559</sup> VLAVLGAVLALLLVLL	ND
CADF_HUMAN	<sup>611</sup> LASALLLVLLVLLVAL	ND
CD72_MOUSE	<sup>93</sup> LQNFLLGLLLSCLMLG	ND
ENV_FRSFB	<sup>339</sup> LLIILLLLILLLWTL	860 ± 205
EPOR_MOUSE	<sup>256</sup> LILVLSILLTVLALL	1170 ± 263
GPBB_HUMAN	<sup>156</sup> LALLGLGLLHALLVL	ND
HEMA_CDVO	<sup>38</sup> LLFVLLILLVGLIALL	770 ± 175
LECH_CHICK	<sup>27</sup> AVYVLLALSFLLLTLL	570 ± 103
PVR_MOUSE	<sup>35</sup> LLVLLLAGGFLALILL	593 ± 142
SRPB_MOUSE	<sup>35</sup> LLSVAVALAVLLTLV	ND
TNRC_MOUSE	<sup>223</sup> LLAILLSLVFLFLLFTT	ND
VE5A_BPVI	<sup>14</sup> AAMQLLLFLFLFFL	600 ± 85
VGLX_HSVBS	<sup>355</sup> LAIALLVLLFSLVIVL	ND

<sup>a</sup> Swiss-Prot sequence identifiers are as follows: CAD1\_XENLA, *Xenopus laevis* E-cadherin; CAD3\_HUMAN, human P-cadherin; CADB\_CHICK, chick B-cadherin; CADF\_HUMAN, human M-cadherin; CD72\_MOUSE, mouse CD72 antigen; ENV\_FRSFB, envelope protein from friend spleen focus-forming virus; EPOR\_MOUSE, mouse erythropoietin receptor; GPBB\_HUMAN, human platelet glycoprotein Ib  $\beta$ -chain; HEMA\_CDVO, hemagglutinin-neuraminidase from canine distemper virus; LECH\_CHICK, chick hepatic lectin; PVR\_MOUSE, mouse poliovirus receptor homolog; SRPB\_MOUSE, mouse signal recognition particle receptor  $\beta$ -subunit; TNRC\_MOUSE, mouse lymphotoxin- $\beta$  receptor; VE5A\_BPVI, E5 protein from bovine papilloma virus; VGLX\_HSVBS, glycoprotein GX from bovine herpesvirus.

<sup>b</sup> Sequences representing those parts of the TMSs that cover the query pattern. The sequence positions of the N-terminal residues are stated, and leucine residues within the heptad pattern given above the sequences are in boldface type.

<sup>c</sup>  $\beta$ -Galactosidase activity as determined with the ToxR system (MU, mean  $\pm$  S.D.).

<sup>d</sup> ND, not determined. The following Swiss-Prot identifiers denote proteins whose signal sequences exhibit the search pattern: A2AP\_BOVIN, AMD\_RAT, AMY\_BACLI, BST1\_HUMAN, C1QC\_HUMAN, C714\_SOLME, CP44\_RABBIT, CP45\_RABBIT, CP46\_RABBIT, CPB1\_RAT, CPB2\_RAT, CPB3\_RAT, CPB4\_RAT, CPB5\_RAT, CPB6\_RAT, CPBA\_MOUSE, CPBB\_CANVA, CPBX\_CAVPO, CPB6\_RAT, CPFL\_HUMAN, CYCH\_PSEFL, CYTN\_HUMAN, ER72\_HUMAN, GDF5\_MOUSE, GVAV\_CAVPO, HEXB\_HUMAN, I12A\_BOVIN, KAIN\_HUMAN, LPL\_BUCAP, PYY\_HUMAN, P2P\_HUMAN, RDHL\_BOVIN, RNL4\_HUMAN, TGFL\_XENLA, THRR\_XENLA, VD15\_RAT, I12A\_PIG, OS9\_HUMAN. The following SWISS-PROT identifiers denote polytopic proteins with the search pattern: ALG3\_YEAST, BTUC\_ECOLI, BVGS\_BORBR, CCKR\_CAVPO, CLC5\_HUMAN, CO3\_RAT, COMT\_HUMAN, COP\_CLOPE, CYB\_BOVIN, DCDR\_XENLA, DHSD\_PORPU, FLO1\_HUMAN, GLR\_MOUSE, HYFF\_ECOLI, I18A\_HUMAN, I18B\_HUMAN, LEP3\_ERWCA, LMP1\_EBU, LMP2\_EBU, NPT2\_HUMAN, NTPL\_ENTHR, NU2M\_CHICK, NU4M\_BRACM, OLIC\_HUMAN, PF2R\_HUMAN, PM22\_MOUSE, PSBC\_MAIZE, ROM1\_BOVIN, TSHR\_HUMAN, VMSA\_HPBGS.

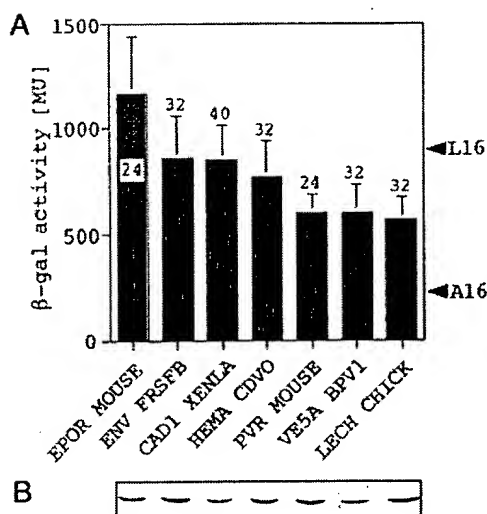
considered in interpreting the results.

A leucine zipper type of side-chain packing also accounts for TMS interactions within phospholamban (7, 8), the M2 proton channel (9), and different polytopic membrane proteins (12). In contrast to our leucine-based model, these heptad motifs are made up of different hydrophobic amino acids, which may generate the characteristically shaped helix surfaces ensuring specific, stoichiometric, and/or heterophilic assembly of these natural proteins.

#### Leucine Zipper Motifs in Natural Membrane Proteins

Data base searching identified leucine-rich heptad motifs within different naturally occurring TMSs, and an analyzed subset of these indeed exhibited various levels of self-interaction. This predicts a role of TMS interactions in the assembly of the corresponding membrane proteins. This is also implied by studies on the corresponding full-length proteins as will be briefly discussed below.

**Cadherins**—Cadherins are calcium-dependent homophilic



**FIG. 4. Transcription activation and expression of ToxR constructs with natural TMSs.** A, transcription activation in FHK12 cells reflects various levels of self-assembly of the TMSs whose amino acid sequences are given in Table 1. The Swiss-Prot identifiers are explained in the legend of Table 1. The bars represent mean specific  $\beta$ -galactosidase activities averaged from 24–32 data points; error bars denote S.D. Arrowheads indicate the signals elicited by the L16 and A16 sequences for comparison (see Fig. 2). B, Western blotting revealed roughly similar expression levels for the different proteins. The order of samples corresponds to that in A.

cell-cell adhesion molecules. Their function depends on lateral clustering within the plasma membrane (34), which is believed to involve interactions between extracellular (35) and juxtamembrane domains (36). On the other hand, leucine-rich heptad motifs are evolutionarily conserved in the TMSs of different cadherin families, and our data demonstrating self-interaction of the E-cadherin TMS suggest a role of TMS interactions in clustering. Strong support for this hypothesis is provided by our recent experimental evidence indicating that mutations reducing the TMS interaction likewise affect the adhesive properties of full-length E-cadherin expressed in eukaryotic cells.<sup>2</sup>

**Erythropoietin Receptor**—The erythropoietin receptor (EpoR) is required for erythrocyte maturation. In analogy to other growth factor receptors, erythropoietin binding is thought to trigger homo-dimerization followed by receptor activation (37). Apart from the case of the Neu oncogene product, where a point mutation within the TMS triggers ligand-independent receptor activation (38), the role of the TMS in growth factor receptor activation is currently not clear. Since ligand binding is translated into activation of cytoplasmic domains, it has been postulated that the subunit-subunit interface of growth factor receptors extends across the membrane and that TMS interactions contribute to ligand-induced subunit assembly in a non-specific way (1). Our finding that the EpoR TMS is capable of self-interaction indeed suggests its contribution to ligand-induced receptor assembly. Alternatively, the EpoR may exist as a preformed dimer activated by ligand binding. Precedence for the latter model is given by the insulin receptor or the aspartate chemoreceptor; in both cases, ligand-binding activates preformed receptor oligomers (39). Ligand-independent dimerization has also been proposed for the epidermal growth factor receptor (40).

**Viral Envelope Proteins**—Enveloped viruses enter the cytoplasm of host cells upon fusion of viral and cellular membranes mediated by fusogenic viral envelope proteins. These proteins

exist as oligomeric complexes (41), and both their fusogenicity and oligomerization appear to depend on their TMSs. For example, the influenza hemagglutinin TMS is required for full membrane fusion (42) and stabilizes the trimeric complex (43). Also, mutations of conserved leucine residues within the TMS of the hemagglutinin-neuraminidase of Newcastle disease virus affected tetramerization and fusion promotion (44). Extending these findings, our data suggest a role of the TMS in oligomerization of hemagglutinin-neuraminidase from canine distemper virus and of the Friend leukemia virus envelope protein. Apart from homooligomerization, a heterophilic and functionally important interaction has been reported between the EpoR and the gp55 protein of Friend spleen focus-forming virus, which is derived from its envelope protein (45). At the surface of infected erythroid cells, the EpoR and gp55 form a noncovalent complex, which results in erythropoietin-independent cell differentiation (46). Complex formation is therefore thought to cause persistent EpoR activation (47). Notably, both the gp55 TMS and the EpoR TMS have been shown to be crucial for this heterophilic interaction (48, 49). Since both TMS sequences have been identified by our data base search and shown to self-interact, we propose that formation of the heteromeric complex proceeds from preformed gp55 and EpoR homomers.

**E5 Protein**—The papillomavirus E5-protein is a transforming membrane protein that exists as a disulfide-bonded dimer (50). Its transforming activity presumably rests on interaction with, and ligand-independent activation of, the receptors for epidermal growth factor, colony-stimulating factor (51), or platelet-derived growth factor (52). In the case of the platelet-derived growth factor receptor, binding to the E5 protein has been directly demonstrated to involve the TMSs plus extracellular flanking regions of both receptor and E5 protein (53). Although the E5 protein extracellular region and the glutamine residue within the TMS are important for activity (54), we speculate that the leucine-rich surface of its TMS aids in homodimer formation and/or binding to the various growth factor receptors TMSs.

**Asialoglycoprotein Receptor**—The hepatic asialoglycoprotein receptors remove abnormally glycosylated proteins from blood circulation (55). The chick homolog exists as a homotrimer whose formation and stability depends on the TMS and flanking sequences (56, 57). This is consistent with self-interaction of its TMS shown here.

These examples demonstrate that assembly of several different natural membrane proteins depends on their TMSs as predicted by the presence of leucine-rich heptad repeats. Future studies will show whether these TMS interactions are based on the leucine zipper type of packing as inferred for our self-assembling model TMSs L16 and AZ2. TMS interactions may be modulated by the lipid composition of the respective host membrane. Further, they may not be the exclusive cause of subunit-subunit recognition but may be complemented by interactions between extramembraneous domains in particular cases.

**Acknowledgments**—We thank Drs. W. Hoch, G. Schönrich, and A. Skerra for critically reading the manuscript; Dr. H. Bedouelle for providing strain PD28; M. Seiler for constructing pSNiR2; A. Szabowski for initial experiments; and Dr. W. B. Huttner for continuous support.

#### REFERENCES

1. Lemmon, M. A., and Engelman, D. M. (1994) *Q. Rev. Biophys.* **27**, 157–218.
2. Dieckmann, G. R., and DeGrado, W. F. (1997) *Curr. Opin. Struct. Biol.* **7**, 486–494.
3. Lemmon, M. A., Flanagan, J. M., Treutlein, H. R., Zhang, J., and Engelman, D. M. (1992) *Biochemistry* **31**, 12719–12725.
4. MacKenzie, K. R., Prestegard, J. H., and Engelman, D. M. (1997) *Science* **276**, 131–133.
5. Laage, R., and Langosch, D. (1997) *Eur. J. Biochem.* **249**, 540–546.

<sup>2</sup> O. Huber, R. Kemler, and D. L., manuscript submitted.

6. Lupas, A. (1996) *Trends Biochem. Sci.* **21**, 375-382
7. Arkin, I. T., Adams, P. D., MacKenzie, K. R., Lemmon, M. A., Brünger, A. T., and Engelman, D. M. (1994) *EMBO J.* **13**, 4757-4764
8. Simmerman, H. K. B., Kobayashi, Y. M., Autry, J. M., and Jones, L. R. (1996) *J. Biol. Chem.* **271**, 5941-5946
9. Pinto, L. H., Dieckmann, G. R., Gandhi, C. S., Papworth, C. G., Braman, J., Shaugnessy, M. A., Lear, J. D., Lamb, R. A., and DeGrado, W. F. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 11301-11306
10. Rees, D. C., Komiya, H., Yeates, T. O., Allen, J. P., and Feher, G. (1989) *Annu. Rev. Biochem.* **58**, 607-633
11. Cohen, C., and Parry, D. A. D. (1990) *Proteins* **7**, 1-15
12. Langosch, D., and Heringa, J. (1998) *Proteins* **31**, 150-160
13. Brosig, B., and Langosch, D. (1998) *Protein Sci.* **7**, 1052-1056
14. Kolmar, H., Hennecke, F., Götze, K., Janzer, B., Vogt, B., Mayer, F., and Fritz, H.-J. (1995) *EMBO J.* **14**, 3895-3904
15. Langosch, D. L., Brosig, B., Kolmar, H., and Fritz, H.-J. (1996) *J. Mol. Biol.* **263**, 525-530
16. Becker, C.-M., Hoch, W., and Betz, H. (1989) *J. Neurochem.* **53**, 124-131
17. Chen, H., and Kendall, D. A. (1995) *J. Biol. Chem.* **270**, 14115-14122
18. Lupas, A., Van Dyke, M., and Stock, J. (1991) *Science* **252**, 1162-1164
19. Schrauber, H., Eisenhaber, F., and Argos, P. (1993) *J. Mol. Biol.* **230**, 592-612
20. Zhang, Y.-P., Lewis, R. N. A. H., Hodges, R. S., and McElhaney, R. N. (1992) *Biochemistry* **31**, 11579-11588
21. Marqusee, S., Robbins, V. H., and Baldwin, R. L. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 5286-5290
22. Serrano, L., Neira, J.-L., Sancho, J., and Fersht, A. R. (1992) *Nature* **356**, 453-455
23. von Heijne, G. (1991) *J. Mol. Biol.* **218**, 499-503
24. Persson, B., and Argos, P. (1994) *J. Mol. Biol.* **237**, 182-192
25. Creighton, T. E. (1993) *Proteins: Structures and Molecular Properties*, pp. 142 and 154, Freeman, New York
26. Bedouelle, H., and Duplay, P. (1988) *Eur. J. Biochem.* **171**, 541-549
27. Lear, J. D., Wasserman, Z. R., and DeGrado, W. F. (1988) *Science* **240**, 1177-1181
28. Whitley, P., Nilsson, I., and von Heijne, G. (1994) *Nat. Struct. Biol.* **1**, 858-862
29. Nezil, F. A., and Bloom, M. (1992) *Biophys. J.* **61**, 1176-1183
30. Webb, R. J., East, J. M., Sharma, R. P., and Lee, A. G. (1998) *Biochemistry* **37**, 673-679
31. Subczynski, W. K., Lewis, R., McElhaney, R. N., Hodges, R. S., Hyde, J. S., and Kusumi, A. (1998) *Biochemistry* **37**, 3156-3164
32. Kuroiwa, T., Sakaguchi, M., Mihara, K., and Omura, T. (1991) *J. Biol. Chem.* **266**, 9251-9255
33. Whitley, P., Grahn, W., Kutay, U., Rapoport, T. A., and von Heijne, G. (1996) *J. Biol. Chem.* **271**, 7583-7586
34. Yap, A. S., Briehar, W. M., Pruschy, M., and Gumbiner, B. M. (1997) *Curr. Biol.* **7**, 308-315
35. Shapiro, L., Fannon, A. M., Kwong, P. D., Thompson, A., Lehmann, M. S., Grubel, G., Legrand, J.-F., Als-Nielsen, J., Colman, D. R., and Hendrickson, W. A. (1995) *Nature* **374**, 327-337
36. Yap, A. S., Niessen, C. M., and Gumbiner, B. M. (1998) *J. Cell Biol.* **141**, 779-789
37. Watowich, S. S., Hilton, D. J., and Lodish, H. F. (1994) *Mol. Cell. Biol.* **14**, 3535-3549
38. Weiner, D. B., Liu, J., Cohen, J. A., Williams, W. V., and Greene, M. I. (1989) *Nature* **339**, 230-231
39. Heldin, C.-H. (1995) *Cell* **80**, 213-223
40. Cadella, T. W. J., and Jovin, T. M. (1995) *J. Cell Biol.* **129**, 1543-1558
41. Hernandez, L. D., Hoffman, L. R., Wolfsberg, T. G., and White, J. M. (1996) *Annu. Rev. Cell Biol.* **12**, 627-661
42. Kemble, G. W., Danieli, T., and White, J. M. (1994) *Cell* **76**, 383-391
43. Doms, R. W., and Helenius, A. (1986) *J. Virol.* **60**, 833-839
44. McGinnes, L., Sergel, T., and Morrison, T. (1993) *Virology* **196**, 101-110
45. Wolff, L., Kaminchik, J., Hankins, W. D., and Ruscetti, S. K. (1985) *J. Virol.* **53**, 570-578
46. Tarr, K., Watowich, S. S., and Longmore, G. D. (1997) *J. Biol. Chem.* **272**, 9099-9107
47. Li, J.-P., D'Andrea, A. D., Lodish, H. F., and Baltimore, D. (1990) *Nature* **343**, 762-764
48. Chung, S.-W., Wolff, L., and Ruscetti, S. K. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 7957-7960
49. Zon, L. I., Moreau, J.-F., Koo, J.-W., Mathey-Prevot, B., and D'Andrea, A. D. (1992) *Mol. Cell. Biol.* **12**, 2949-2957
50. Horwitz, B. H., Burkhardt, A. L., Schlegel, R., and DiMaio, D. (1988) *Mol. Cell. Biol.* **8**, 4071-4078
51. Martin, P., Vass, W. C., Schiller, J. T., Lowy, D. R., and Velu, T. J. (1989) *Cell* **59**, 21-32
52. Petti, L., and DiMaio, D. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 6736-6740
53. Petti, L. M., Reddy, V., Smith, S. O., and DiMaio, D. (1997) *J. Virol.* **71**, 7318-7327
54. Kulke, R., Horwitz, B. H., Zibello, T., and DiMaio, D. (1992) *J. Virol.* **66**, 505-511
55. Drickamer, K. (1995) *Curr. Opin. Struct. Biol.* **5**, 612-616
56. Loeb, J. A., and Drickamer, K. (1987) *J. Biol. Chem.* **262**, 3022-3029
57. Verrey, F., and Drickamer, K. (1993) *Biochem. J.* **292**, 149-155

## Validation of the single-stranded channel conformation of gramicidin A by solid-state NMR

F. KOVACS\*†, J. QUINE\*‡, AND T. A. CROSS\*†§¶

\*National High Magnetic Field Laboratory, †Institute of Molecular Biophysics, and Departments of ‡Chemistry and §Mathematics, Florida State University, Tallahassee, FL 32306-4005

Communicated by Alexander Pines, University of California, Berkeley, CA, May 14, 1999 (received for review February 11, 1999)

**ABSTRACT** The monovalent cation selective channel formed by a dimer of the polypeptide gramicidin A has a single-stranded, right-handed helical motif with 6.5 residues per turn forming a 4-Å diameter pore. The structure has been refined to high resolution against 120 orientational constraints obtained from samples in a liquid-crystalline phase lipid bilayer. These structural constraints from solid-state NMR reflect the orientation of spin interaction tensors with respect to a unique molecular axis. Because these tensors are fixed in the molecular frame and because the samples are uniformly aligned with respect to the magnetic field of the NMR spectrometer, each constraint restricts the orientation of internuclear vectors with respect to the laboratory frame of reference. The structural motif of this channel has been validated, and the high-resolution structure has led to precise models for cation binding, cation selectivity, and cation conductance efficiency. The structure is consistent with the electrophysiological data and numerous biophysical studies. Contrary to a recent claim [Burkhart, B. M., Li, N., Langs, D. A., Pangborn, W. A. & Duax, W. L. (1998) *Proc. Natl. Acad. Sci. USA* 95, 12950–12955], the solid-state NMR constraints for gramicidin A in a lipid bilayer are not consistent with an x-ray crystallographic structure for gramicidin having a double-stranded, right-handed helix with 7.2 residues per turn.

Oriental constraints derived from solid-state NMR can be used to determine high-resolution three-dimensional structures. Such an approach has been used to define the structure of the ion channel, gramicidin A, in lamellar phase lipids (ref. 1; PDB accession no. 1MAG). Although a reasonable model of this structure has been extant for nearly 30 years (2) and a structure was determined by solution NMR spectroscopy in SDS micelles (3, 4), crystallographic and solution NMR methods have not been successful in a lipid environment. Recently, the validity of the solid-state NMR structure has been challenged (5). In this report, the structural fold of the channel is validated by comparing predicted and observed values for structural constraints not used quantitatively in solving for the structural fold. Furthermore, the NMR observables are compared with predicted values from several structures in the Protein Data Bank. The results establish the high resolution of the solid-state structure and the clear validity of this motif in a lipid environment.

Gramicidin A is a polymorphic structure and the dominant sequence in gramicidin D, the biosynthetic product from *Bacillus brevis*: HCO-Val-Gly-Ala-DLeu-Ala-DVal-Val-Val-Trp-DLeu-Trp-DLeu-Trp-DLeu-Trp-NHCH<sub>2</sub>CH<sub>2</sub>OH. In isotropic organic solvents, this peptide typically forms a double-stranded dimer that may be parallel or antiparallel, left-handed or right-handed and has a range of residues per turn from 5.6 to 7.2 (5–10). In the heterogeneous anisotropic lipid

environment, the structure is almost exclusively single-stranded. Because of the short helix length, a single-stranded monomer buries one of its termini in the bilayer. Exposure of the carboxyl terminus to the bilayer surface and lack of exposure for the amino terminus was documented by shift reagent NMR experiments (11). It has been observed that the native monovalent cation selective channel function is maintained when formal charges are introduced at the carboxyl terminus but not when they are introduced at the amino terminus (12). Circular dichroism can distinguish between single-stranded and double-stranded conformers (13, 14), and bilayer preparations of gramicidin have been shown, based on this technique, to be single-stranded. Low-angle x-ray scattering of bilayer preparations has characterized the helical pitch as single-stranded, not double-stranded (15). Moreover, the structure in the membrane-mimetic SDS micelles is definitively single-stranded, and the solid-state NMR structure described here in lamellar phase lipids is also definitively single-stranded. The only structurally characterized double-stranded conformer of gramicidin A in a lipid bilayer was shown to be in a kinetically trapped state that, on heating at 68°C for 3 days, was converted to the single-stranded channel state (16).

The helical sense of the channel state was originally thought to be left-handed (17), but both the SDS micellar structure (3) and the solid-state NMR data (18) in hydrated lipid bilayers clearly showed that the gramicidin A conformation in membrane mimetic environments was right-handed.

**Structural Determination from Orientational Constraints.** Oriental constraints are derived from the anisotropic nuclear spin interactions observed by solid-state NMR of uniformly aligned samples. Isotopic labeling has been achieved by using solid-phase peptide synthesis with Fmoc blocking chemistry and HPLC purification when necessary, but typical purity before HPLC purification is >95% (19, 20). As shown by NMR, alignment of the samples with a mosaic spread as small as 0.3° (21) has been obtained for gramicidin A in dimyristoyl phosphatidylcholine bilayers (1:8 molar ratio and ≈50% by weight water). The small mosaic spread has been achieved by preparing samples on thin glass slides, and, in the NMR magnet, the peptides, through their diamagnetic susceptibility, have a tendency to align parallel to the magnetic field axis.

Observed dipolar splittings, such as <sup>15</sup>N–<sup>1</sup>H, <sup>15</sup>N–<sup>13</sup>C, and <sup>15</sup>N–<sup>2</sup>H have a cos<sup>2</sup>θ dependence with respect to the magnetic field axis. The observed dipolar splitting can be interpreted in light of the magnitude of the dipolar coupling,  $\nu_d$ . This magnitude depends on physical constants, the distance separating the two nuclei, and a characterization of motional averaging. The motional averaging can be assessed independently and has been for gramicidin throughout the molecule (22–24). This interpretation of the observed dipolar splitting,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviation: CHARMM, Chemistry at Harvard Macromolecular Mechanics.

¶To whom reprint requests should be addressed. e-mail: cross@magnet.fsu.edu.

therefore, leads directly to the orientation of the internuclear vector with an error that is dominated by the error in the dipolar observations.

The dipolar and  $^2\text{H}$  quadrupolar interactions are essentially axially symmetric interactions, and hence, their magnitude is characterized by a single number. The anisotropic chemical shift is an axially asymmetric interaction characterized by three tensor elements whose magnitudes can be assumed along with a substantial error bar or experimentally characterized from the observation of an unoriented powder pattern. Experimental characterization has been done for each of the  $^{15}\text{N}$  chemical shift tensors in the gramicidin backbone (25). In addition, the orientations of approximately half of the chemical shift tensors have been characterized with respect to the molecular frame (25, 26). Based on these characterizations, reasonable assumptions were made about the tensor orientation for the other sites. Therefore, the error for chemical shifts reflects not only the error in observation of the anisotropic chemical shift but also a small contribution from tensor characterization.

The interpretation of dipolar and quadrupolar data,  $\Delta\nu_{\text{obs}}$ , as orientational constraints leads to several ambiguities in the orientation of the internuclear vectors, represented by unit vectors,  $u$ .

$$\Delta\nu_{\text{obs}} = \nu_{\parallel}[3(B \cdot u)^2 - 1] \quad [1]$$

$B$  is a unit vector in the direction of the magnetic field and  $B \cdot u$  has values between  $-1$  and  $1$ . Therefore,  $\Delta\nu_{\text{obs}}$  is positive if  $\Delta\nu_{\text{obs}} > \nu_{\parallel}$ , but if  $\Delta\nu_{\text{obs}} \leq \nu_{\parallel}$ , it can be either positive or negative, a sign that is not readily obtained by experiment. Even when  $\Delta\nu_{\text{obs}}$  is positive,  $B \cdot u$  can be either positive or negative. Hence, there are either two or four possible orientations for each internuclear vector consistent with the dipolar constraint. The interpretation of spin interactions with axially asymmetric tensors, such as the chemical shift, does not give rise to discrete solutions for the internuclear vectors,

$$\sigma_{\text{obs}} = \sigma_{11}(B \cdot \sigma_{11})^2 + \sigma_{22}(B \cdot \sigma_{22})^2 + \sigma_{33}(B \cdot \sigma_{33})^2, \quad [2]$$

where  $\sigma_{ii}$  are unit vectors that define the orientation of the chemical shift tensor elements and  $\sigma_{\text{obs}}$  is the chemical shift in an aligned sample. Because

$$(B \cdot \sigma_{11})^2 + (B \cdot \sigma_{22})^2 + (B \cdot \sigma_{33})^2 = 1, \quad [3]$$

there is not a unique solution or even a discrete set of solutions to these two equations with three unknowns. For  $^{15}\text{N}$  chemical shift tensors, two tensor elements,  $\sigma_{11}$  and  $\sigma_{33}$ , are typically in the peptide plane. Therefore,  $\sigma_{11}$  and  $\sigma_{33}$  can be rewritten as

$$\sigma_{11} = Au_1 + Bu_2 \text{ and } \sigma_{33} = Cu_1 + Du_2, \quad [4]$$

where  $A$ ,  $B$ ,  $C$ , and  $D$  are known from the covalent geometry and the orientation of the tensor relative to the molecular frame. If  $u_1$  and  $u_2$  are chosen, such that  $B \cdot u_1$  and  $B \cdot u_2$  are known, albeit with ambiguity, from dipolar or quadrupolar interactions, then the use of

$$\begin{aligned} \sigma_{\text{obs}} - \sigma_{22} &= (\sigma_{11} - \sigma_{22})(Au_1 \cdot B + Bu_2 \cdot B)^2 \\ &+ (\sigma_{33} - \sigma_{22})(Cu_1 \cdot B + Du_2 \cdot B)^2 \end{aligned} \quad [5]$$

will greatly reduce the number of possible orientations for the peptide plane containing  $u_1$  and  $u_2$  (Fig. 1).

In gramicidin for an isolated peptide plane, two solutions remain, identical orientations with respect to the  $+B$  and  $-B$  directions. This ambiguity is not a problem for two reasons. First, this ambiguity reflects whether the molecule as a whole is oriented with respect to the  $+B$  or  $-B$  field directions. Because both molecular orientations exist in our aligned samples and because the NMR observables are independent of the sign, this ambiguity is not a problem. However, when considering the adjacent peptide plane, this ambiguity affects the relative orientation of the peptide planes and hence the conformation. Fortunately, the relative orientation is typically defined through a combination of the  $\text{C}_\alpha\text{-}^2\text{H}$  orientational constraint and the covalent geometry surrounding the  $\text{C}_\alpha$  site.

Although a unique set of bond orientations is hereby defined with respect to  $B$ , a final ambiguity remains, the orientation of the normal to the peptide plane,  $B \cdot u_1 \times u_2$ . The sign of this triple product is not defined. Because the angle of  $B$  to the peptide plane is small and because the  $\text{C}_\alpha\text{-C}_\alpha$  axis is nearly perpendicular to  $B$ , the sign of this triple product has a very small effect on the position of the  $\text{C}_\alpha$  carbons. Consequently, this ambiguity has little effect on the helical parameters and on the determination of the molecular fold. As will be shown later, this ambiguity, known as a chirality ambiguity, can be resolved in the refinement procedure for the molecular structure. The unique nature of the molecular fold has been illustrated by assembling four initial structures (Fig. 1B) with differing

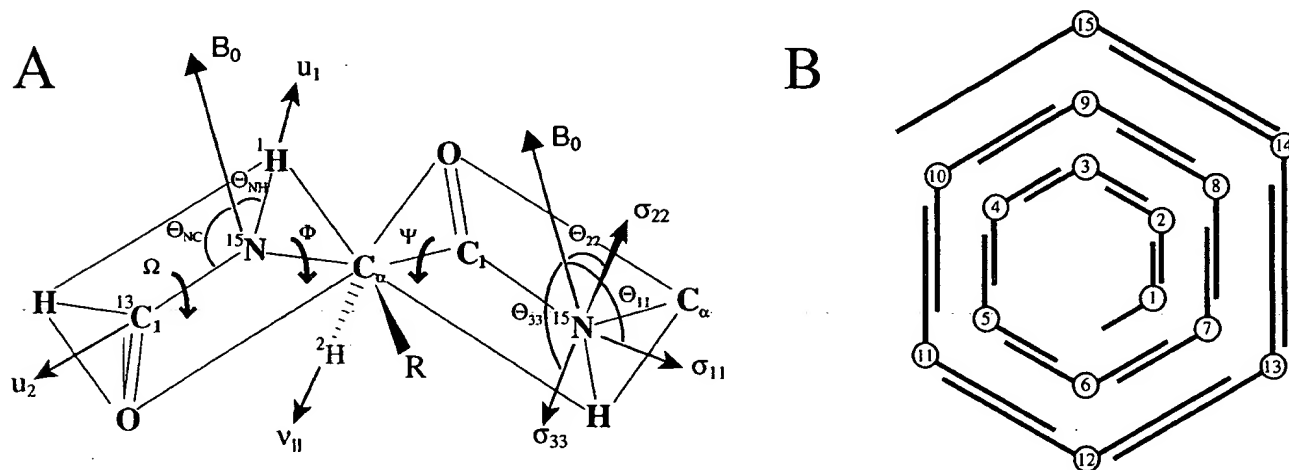


FIG. 1. (A) The primary structural constraints for the polypeptide backbone are derived from the  $^{15}\text{N}\text{-}^1\text{H}$  and  $^{15}\text{N}\text{-}^{13}\text{C}$  dipolar interactions, the  $\text{C}_\alpha\text{-}^2\text{H}$  quadrupolar interaction, and the anisotropic  $^{15}\text{N}$  chemical shift. The initial structure is developed by determining each peptide-plane orientation with respect to the magnetic field axis with two dipolar interactions. The relative orientations of the peptide planes is then determined (i.e., the  $\phi$  and  $\psi$  angles) for a diplane structure. (B) The initial structure is assembled sequentially with overlapping diplanes. The initial structure is not a unique structure because of chirality ambiguities; however, the molecular fold, hydrogen-bonding pattern, helical sense, and residues per turn are uniquely defined.



chirality patterns: (i) all chiralities positive, (ii) all chiralities negative, (iii) chiralities alternating  $+/-$ , and (iv) chiralities alternating  $-/+$  (24). The four structures all have the same helical sense, number of residues per turn, and hydrogen-bond pattern. This gramicidin A channel conformation is a  $\beta$ -strand in which, because of the alternating pattern of D and L amino acid stereochemistry, all of the side chains radiate toward the lipid environment, leaving the polypeptide backbone to line the channel. The 6.5 residues per turn result in a nominal channel diameter of 4 Å, which permits just a single-file column of water molecules. There are 10 parallel  $\beta$ -sheet-type hydrogen bonds stabilizing this monomeric structure. In addition, six intermolecular hydrogen bonds stabilize the dimer at the bilayer center through antiparallel  $\beta$ -sheet-type hydrogen bonds. Although the monomeric unit of the channel structure has been characterized previously, only recently has there been direct solid-state NMR distance measurements confirming this model of the monomer-monomer interface (R. Fu, M. Cotten, and T.A.C., unpublished results).

Indeed, in building these structures several important features of the orientational constraints are illustrated. The initial structures have hydrogen-bond distances within an rms deviation of 0.5 Å of the ideal  $\beta$ -sheet hydrogen bonds (27). The orientational constraints must be both precise and accurate, because 14 separate dipolar orientational constraints are used quantitatively to define a turn of the helix. Furthermore, given the initial assumption of Engh and Huber (28) covalent geometry and 180°  $\omega$ -torsion angles seem very reasonable. Finally, errors associated with each orientational constraint, even errors of only a couple of degrees, would result in far worse hydrogen-bond distances if the errors accumulated; however, because the constraints fix each site independently with respect to the laboratory  $z$  axis, the errors do not sum. In other words, the orientational constraints are absolute as opposed to relative constraints.

**Structural Refinement.** The orientational constraints have not been used optimally in the development of the initial structures. Although the dipolar constraints have been used quantitatively, the chemical shifts and the  $C_{\alpha}$ - $^2H$  quadrupolar constraints have been used only as filters to eliminate certain possible peptide-plane orientations. In a refinement protocol, the structure will be refined against a generalized global penalty function including all of the orientational constraints, as well as ideal hydrogen-bond geometry and the Chemistry at Harvard Macromolecular Mechanics (CHARMM) force field. The refined structure is obtained through a geometrical search in which the NMR observables and conformational parameters are calculated for each structural modification and compared with the observed data, ideal hydrogen-bond geometry, and the CHARMM energy of the previous structure. The conformational search and evaluation is particularly difficult with the accurate orientational constraints. The possible conformations are separated by very high-penalty barriers, and therefore, an adequate search of the conformational space required a different approach. Three types of structural modifications were implemented: (i) random atom moves with a diffusion parameter of  $5 \times 10^{-4}$  Å; (ii) compensating torsional moves for  $\psi_i$  and  $\phi_i + 1$  of equal magnitude ( $\leq 3^\circ$ ) and opposite sign; and (iii) tunneling moves, a specialized form of compensating torsional moves, designed to approximate a change in chirality. Simulated annealing was used to perform the minimization of the penalty function (29) and to generate a structure with minimized energy and optimized fit to the experimental data. Moreover, initial assumptions, such as uniform covalent geometry and  $\omega = 180^\circ$ , were relaxed.

For this refinement, the experimental data were weighed heavily compared with the CHARMM force field energy, because the experimental constraints were obtained from samples within a lipid bilayer environment, whereas the CHARMM energy was calculated in the absence of both water

and lipid. The balance of the contributions to the penalty function represented a difficult choice between accurate experimental constraints and an important force field used to maintain appropriate covalent geometry and van der Waals contacts. Actually, a few significant distortions in bond angles have been identified by PROCHECK (1), indicating that further development of the refinement protocol is needed.

In refining the four initial structures with differing chiralities, a unique chirality solution was achieved for nearly all of the peptide planes. The rms deviation between all 40 refinements was just 0.48 Å. To achieve the structure deposited in the Protein Data Bank (Fig. 24), these 40 structures were averaged, and a final refinement was performed by using only atom moves and not torsional moves in the simulated annealing.

**Validation of the Structural Fold.** Although there are opportunities to modify and potentially improve the refinement protocol, the solid-state NMR structure in lamellar phase lipids is a high-resolution structure precisely constrained by the orientational constraints. Cross-validation of solution NMR and x-ray crystallographic structures can be achieved by leaving some of the data out of the structure determination followed by a comparison of predicted values from the resultant structure and the observed values for the data not used in the structure determination. Because of the more limited number of constraints in solid-state NMR, the opportunities for cross-validation are less. However, the initial backbone structure determination was achieved with just the quantitative use of the dipolar constraints. A calculation of the chemical shifts from the initial structure and comparison to the experimental data generate a penalty contribution of less than one error bar per constraint. Moreover, the  $C_{\alpha}$ - $^2H$  quadrupolar splittings lend additional validity to this fold as will be discussed later with Fig. 3. The solid-state NMR structure is both a high-resolution structure and an accurate structure. Furthermore, the experimental constraints have been obtained from samples of gramicidin A solubilized in liquid crystalline phase lipid bilayers. Because this environment is dynamic, it is important to recognize that this structural solution is a time-averaged structure, for which the characterized molecular motions have been taken into consideration through averaging of the nuclear spin interaction tensors.

**Gramicidin Structural Polymorphism.** Recently, Duax and coworkers (5) published a crystal structure of gramicidin, a right-handed, antiparallel, double-stranded structure with 7.2 residues per turn. The authors mistakenly claim that their structure agrees with  $^{15}N$ -NMR data on the functional gramicidin D channel in lipid bilayers (5) and that the solid-state NMR characterized structure "does not have an open channel for ion passage" (5).

The referenced solid-state NMR data were resonances from a chemical shift spectrum of uniformly  $^{15}N$ -labeled gramicidin D in oriented dimyristoyl phosphatidylcholine bilayers (30). Resonance assignments and single-site resolution were not available in 1987, although they have been for the past 5 years. To claim agreement with the NMR data, the authors fabricate a new linear scale in  $^oN-H$ , the angle between the N-H internuclear axis and  $B$ , drawn parallel to the chemical shift scale. Only with significant assumptions, as described in Nicholson *et al.* (30), is the anisotropic chemical shift proportional to  $\cos^2\theta$ , and never is the chemical shift linearly proportional to  $\theta$ . Furthermore, this scale for  $\theta$  from 0 to 90° is displayed over a 230-ppm chemical shift range, rather than the maximum amide  $^{15}N$  chemical shift anisotropy in gramicidin A of 170 ppm. On their scale, the authors have presented values of  $^oN-H$  from their structure and from the single-stranded structure defined by solid-state NMR for comparison to the chemical shift.

Here, we have redone this invalid analysis by accurately predicting the NMR observables for four different gramicidin structures (Fig. 2). These represent structures determined by solution



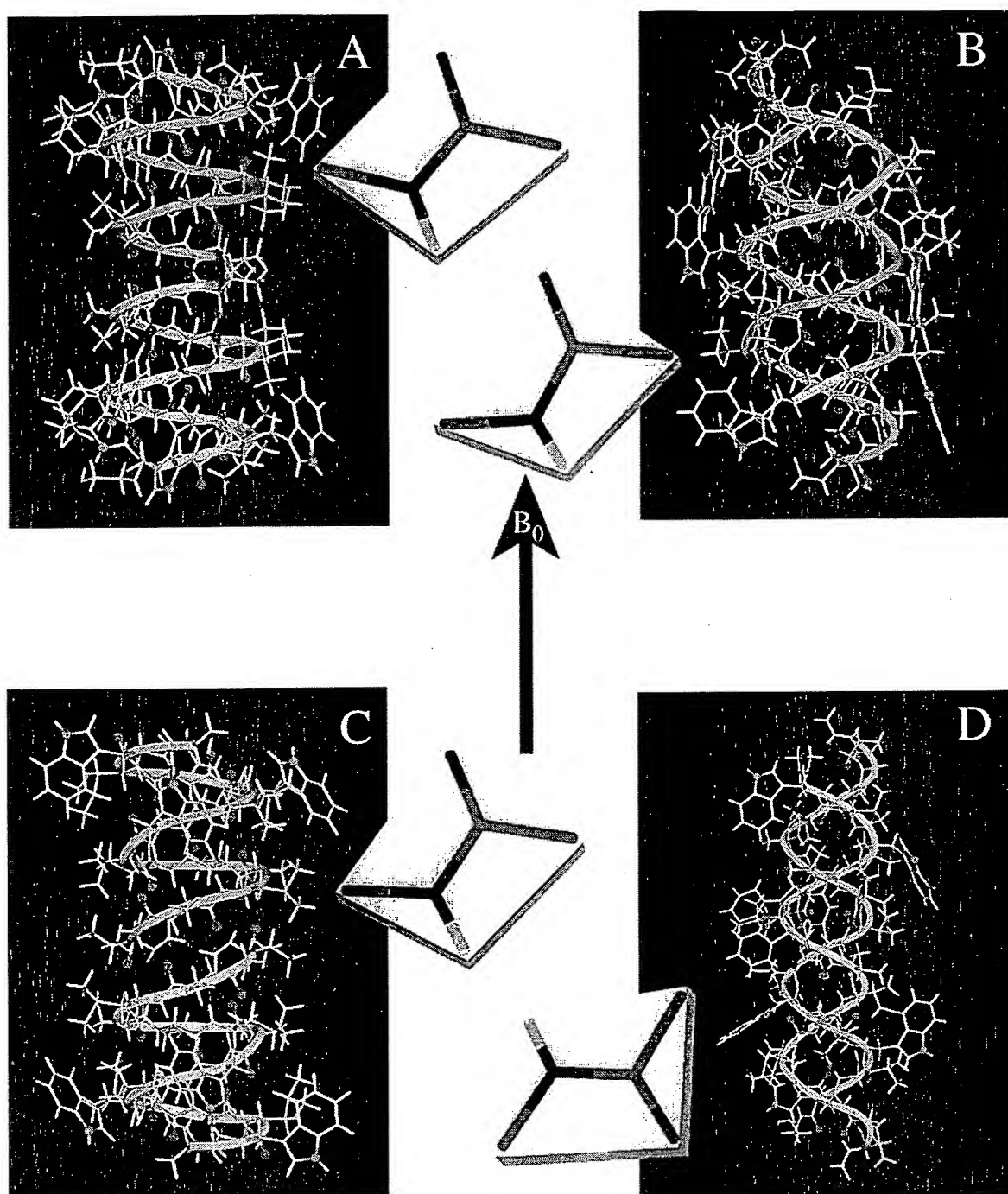


FIG. 2. Gramicidin A structures in different environments. In addition to the atomic ball-and-stick structures, a ribbon is added to accentuate the handedness and strandedness of the helix. The two monomers have different colored ribbons: one yellow and one orange. The backbone carbonyl oxygens that line the pore of the channel are highlighted in red, and the indole  $^{15}\text{N}$  sites, important in dictating the strandedness of the structure in a membrane environment, is shown in blue. For each structure the  $\text{Ala}^3\text{-Leu}^4$  peptide-plane orientation is shown with respect to *B*. (A) The solid-state NMR-derived structure from a bilayer environment: single-stranded, right-handed, and 6.5 residues per turn (ref. 1; PDB accession no. 1MAG). (B) An x-ray crystallographic structure of crystals prepared from  $\text{Cs}^+/\text{MeOH}$  solution: double-stranded, right-handed, and 7.2 residues per turn (ref. 5; PDB accession no. 1AV2). (C) A solution NMR structure from an SDS micellar environment: single-stranded, right-handed, and 6.3 residues per turn (ref. 4; PDB accession no. 1GRM). (D) An x-ray crystallographic structure of crystals prepared from benzene/methanol solution: double-stranded, left-handed, and 5.6 residues per turn (ref. 7; PDB accession no. 1ALZ).

NMR (Fig. 2C), x-ray crystallography (Fig. 2B and D), and solid-state NMR (Fig. 2A). They also represent right-handed (Fig. 2A–C) and left-handed (Fig. 2D) helices, as well as single-stranded (Fig. 2A and C) and double-stranded (Fig. 2B and D) helices. Although these are major differences and the residues per turn vary from 5.6 to 7.2, the secondary structure is

a  $\beta$ -strand for all of them. The predicted values of orientational constraints are directly compared by using a normalized difference between predicted and observed values. Normalization was achieved by dividing the differences by the observed error-bar magnitude. These results are presented in Fig. 3 as a histogram for each structure and organized by constraint type.

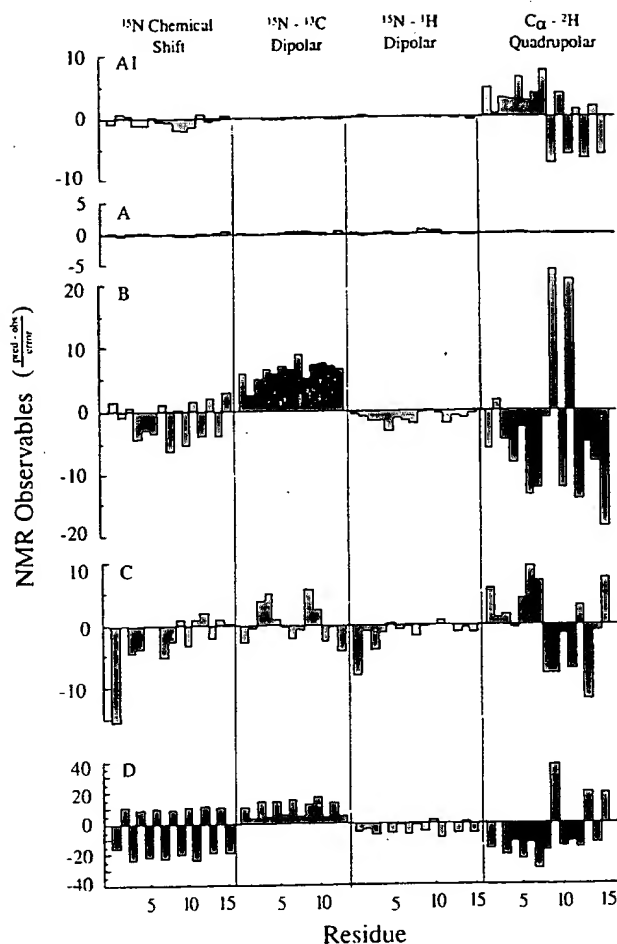


FIG. 3. Predicted NMR observables from the four structures shown in Fig. 2 are compared with the experimental values. The vertical scale is the difference in predicted and observed values of the NMR observables for the backbone divided by the error bar for each class of observables: 5 ppm for the  $^{15}\text{N}$  chemical shift, 100 Hz for the  $^{15}\text{N}$ - $^{13}\text{C}$  dipolar interaction; 2 kHz for the  $^{15}\text{N}$ - $^1\text{H}$  dipolar interaction; and 5 kHz for the  $\text{C}_\alpha$ - $^2\text{H}$  quadrupolar interaction. In addition to the deviations for the four structures shown in Fig. 2 (letters A-D here correspond to structures A-D in Fig. 2), the deviations are also shown for the solid-state NMR derived initial structure (A1). This structure is calculated based on the dipolar constraints and not on the  $^{15}\text{N}$  chemical shift or  $\text{C}_\alpha$ - $^2\text{H}$  quadrupolar interactions; therefore, the deviations displayed for these data represent a validation of the initial structure that defines the structural motif.

Clearly, the structure developed from and refined against the solid-state NMR data (Fig. 2A) is most consistent with the observed data (Fig. 3A). Among the other three structures, the next best fit to the solid-state NMR data is the Arseniev structure from SDS micelles (Figs. 2C and 3C) that has the same fold. Although there are very significant deviations from the observed data, there is no consistent pattern of deviation as there is for both of the double-stranded structures, suggesting that the fold is correct but that details of the peptide-plane orientations tilting into and away from the channel axis are in error. Indeed, such characterizations are beyond the resolution of this solution NMR structure. Note that the scale for the analysis of the 1ALZ structure (Fig. 3D) has been compressed by a factor of four to save space; this structure is the most inconsistent with the solid-state NMR data. The alternating sign pattern in the chemical shift deviations is the result of the opposite helical sense for this structure. Because these structures are all  $\beta$ -strand-type structures, the repeat unit is a dipeptide, and the orientations of the two planes are quite

different. A change in handedness results in an inversion of the peptide planes as shown in Fig. 2. The systematic error in the  $^{15}\text{N}$ - $^{13}\text{C}$  dipolar data for both of the double-stranded structures is indicative of an error in the helical pitch. The axis of the N-C $_1$  bond in the peptide backbone is very sensitive to the helical pitch. Consequently, the correct fold of the solution NMR structure has an average deviation from the observed  $^{15}\text{N}$ - $^{13}\text{C}$  dipolar data that is less than 1 (in units of the error bar; Fig. 3C). The  $\text{C}_\alpha$ - $^2\text{H}$  deviations are some of the largest, but they are also the most precise constraints; the error bar used to normalize these deviations is small compared with the magnitudes of the observables. Furthermore, these data are very sensitive to the angle formed between peptide planes, resulting from a combination of the helical pitch and residues per turn. The average deviations from the solution NMR structure (Fig. 3C) nearly cancel, but the deviations for the double-stranded structures (Fig. 3B and D) have a very significant average value, indicative of systematic errors.

Fig. 3A1 shows the deviations for one of the initial solid-state NMR structures. Because the initial structures differ only by chirality ambiguities, the deviations between predicted and observed data are the same for all of the initial structures. The deviations for the  $^{15}\text{N}$  chemical shift and the  $\text{C}_\alpha$ - $^2\text{H}$  quadrupolar splittings are even smaller than those for the solution NMR structure. These data were not used in the calculation of the initial structure, and consequently, they represent a validation of the initial structure and hence the fold of the polypeptide in lipid bilayers.

It is also possible to refine the other structures against the NMR observables. In refining the 1AV2 structure (Fig. 2B) with full atom and torsional moves, 22 of the 30 N-C $_1$ -C $_a$  and C $_1$ -N-C $_a$  bond angles are three or more standard deviations removed from the Engh and Huber (28) geometry as assessed by PROCHECK, as the refinement protocol attempts to change the N-C $_1$  bond orientation to be consistent with the experimental data. Furthermore, if atom moves are inactivated so that the covalent geometry remains fixed and if the hydrogen-bond distances are inactivated in one of the two grooves of the double-stranded structure, then the number of residues per turn changes from 7.2 to  $\approx 10$  to accommodate the N-C orientations. Finally, if all hydrogen-bond distances are constrained, the experimental data are not well fit, and the structure is still substantially distorted ( $\gg 1$ -Å rms deviation) with respect to the crystallographic coordinates. Moreover, the initial solid-state NMR structure and the solution NMR structure can be refined readily against the experimental data to a good fit even when atom moves are turned off.

The primary argument presented by Duax and coworkers (5) for claiming that their structure is the membrane active form is that their structure was consistent with the solid-state NMR data obtained in a bilayer environment. Because this statement is inaccurate, there is no reason to think that this structure crystallized from methanol solution is the channel conformation that occurs in lipid bilayers. Moreover, it is well understood why the membrane active form is single-stranded. Indoles are much more stable in the hydrophilic/hydrophobic bilayer interface than in the hydrophobic core of the bilayer (21, 31-33). When the tryptophans are completely replaced by phenylalanines, the predominant conformation in the bilayer is the left-handed, double-stranded structure, the same fold as shown in Fig. 2D (34, 35). The indoles are distributed along the molecular axis in both double-stranded structures (Fig. 2B and D) as opposed to the interface location in both single-stranded structures. Therefore, it has been argued that the tryptophan's propensity for the bilayer interface is the primary reason for the structural conversion from double-stranded to single-stranded (35). Clearly, the polypeptide environment is an important factor in dictating the molecular fold of this structure. Consequently, modeling this very heterogeneous bilayer

environment with a homogeneous model, such as an isotropic organic solvent, may not be adequate.

Although the structure described by Duax and coworkers (5) is not the membrane active form, it does illustrate the polymorphic structural nature of this molecule, reflecting the numerous environments in which it has been studied. Indeed, this polymorphism provides a clear example that the amino acid sequence is not the sole determinant of conformation. For instance, a correlation between the antiparallel double-stranded conformation and nonpolar organic solvents has been established (36). In more polar organic solvents, the parallel double-stranded conformation with a net axial dipole is much more stable. It should be noted that this "new fold" was previously described by solution NMR (37).

Gramicidin A has proven to be an excellent model channel for understanding cation selectivity and conductance efficiency. The cation binding site in the channel conformation has been shown to include three or more water ligands (38). Such flexibility results in only modest selectivity among monovalent cations, as the binding site can accommodate cations of various size. Delocalized cation binding leads to a shallow potential-energy well, a minimized entropic penalty for cation binding, and a stepwise dehydration mechanism leading to high cation-association rates. In the very extensive literature from electrophysiologists, molecular dynamicists, and other biophysicists, gramicidin has developed into a great tool for understanding cation conductance. There is little question that the cation conducting conformation for gramicidin is the single-stranded, right-handed structure with 6.5 residues per turn and a 4-Å pore that supports a single-file column of water molecules and monovalent cation transport across membranes.

Moreover, it is shown here that solid-state NMR-derived orientational constraints can lead to both a precise and accurate high-resolution structure in an environment that requires neither isotropic solution nor crystallization. Indeed, this approach has many advantages for characterizing structures in liquid crystalline lipid bilayer environments.

This work was supported by National Institutes of Health Grant AI-23007 and National Science Foundation Grant MCB 9603935. The work was largely performed at the National High Magnetic Field Laboratory supported by the National Science Foundation Cooperative Agreement DMR-9527035 and the State of Florida.

- Ketchum, R. R., Roux, B. & Cross, T. A. (1997) *Structure (London)* **5**, 1655–1669.
- Urry, D. W. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 672–676.
- Bystrov, V. F., Arseniev, A. S., Barsukov, I. L. & Lomize, A. L. (1987) *Bull. Magn. Reson.* **8**, 84–94.
- Lomize, A. L., Orekhov, V. & Arseniev, A. S. (1992) *Bioorg. Khim.* **18**, 182–200.
- Burkhart, B. M., Li, N., Langs, D. A., Pangborn, W. A. & Duax, W. L. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 12950–12955.
- Veatch, W. R., Fossel, E. T. & Blout, E. R. (1974) *Biochemistry* **13**, 5249–5256.
- Langs, D. A. (1988) *Science* **241**, 188–191.
- Arseniev, A. S., Bystrov, V. F., Ivnov, V. T. & Ovchinnikov, Y. A. (1984) *FEBS Lett.* **165**, 51–56.
- Pascal, S. M. & Cross, T. A. (1992) *J. Mol. Biol.* **226**, 1101–1109.
- Doyle, D. A. & Wallace, B. A. (1997) *J. Mol. Biol.* **266**, 963–977.
- Weinstein, S., Wallace, B. A., Blout, E. R., Morrow, J. S. & Veatch, W. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4230–4234.
- Bamberg, E., Apell, H. J. & Alpes, H. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 2402–2406.
- Wallace, B. A., Veatch, W. R. & Bout, E. R. (1981) *Biochemistry* **20**, 5754–5760.
- Bano, M. C., Braco, L. & Abad, C. (1989) *FEBS Lett.* **250**, 67–71.
- Katsaras, J., Prosser, R. S., Stinson, R. H. & Davis, J. H. (1992) *Biophys. J.* **61**, 827–830.
- Arumugam, S., Pascal, S., North, C. L., Hu, W., Lee, K. C., Cotten, M., Ketchum, R. R., Xu, F., Brennenman, M. & Kovacs, F. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 5872–5876.
- Urry, D. W., Walker, J. T. & Trapane, T. L. (1982) *J. Membr. Biol.* **69**, 225–231.
- Nicholson, L. K. & Cross, T. A. (1989) *Biochemistry* **28**, 9379–9385.
- Fields, C. G., Fields, G. B., Noble, R. L. & Cross, T. A. (1989) *Int. J. Pept. Protein Res.* **33**, 298–303.
- Fields, G. B., Fields, C. G., Petefish, J., Van Wart, H. E. & Cross, T. A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1384–1388.
- Hu, W., Lazo, N. D. & Cross, T. A. (1995) *Biochemistry* **34**, 14138–14146.
- North, C. L. & Cross, T. A. (1995) *Biochemistry* **34**, 5883–5895.
- Lee, K.-C., Huo, S. & Cross, T. A. (1995) *Biochemistry* **34**, 857–867.
- Ketchum, R. R., Lee, K.-C., Huo, S. & Cross, T. A. (1996) *J. Biomol. NMR* **8**, 1–14.
- Mai, W., Hu, W., Wang, C. & Cross, T. A. (1993) *Protein Sci.* **2**, 532–542.
- Teng, Q., Iqbal, M. & Cross, T. A. (1992) *J. Am. Chem. Soc.* **114**, 5312–5321.
- Jeffrey, G. A. & Saenger, W. (1994) *Hydrogen Bonding in Biological Systems* (Springer, Berlin).
- Engh, R. A. & Huber, R. (1991) *Acta Crystallogr. A* **47**, 392–400.
- Kirkpatrick, S., Gelatt, C. D., Jr., & Vecchi, M. P. (1983) *Science* **220**, 671–680.
- Nicholson, L. K., Moll, F., Mixon, T. E., LoGrasso, P. V., Lay, J. C. & Cross, T. A. (1987) *Biochemistry* **26**, 6621–6626.
- O'Connell, A. M., Koeppe, R. E., II, & Andersen, O. S. (1990) *Science* **250**, 1256–1259.
- Jacobs, R. E. & White, S. H. (1989) *Biochemistry* **28**, 3421–3437.
- Michel, M. & Deisenhofer, J. (1990) *Curr. Top. Membr. Transp.* **36**, 53–69.
- Cotten, M., Xu, F. & Cross, T. A. (1997) *Biophys. J.* **73**, 614–623.
- Cotten, M., Fu, R. & Cross, T. A. (1999) *Biophys. J.* **76**, 1179–1189.
- Xu, F. & Cross, T. A. (1998) *Magn. Reson. Chem.* **36**, 651–655.
- Arseniev, A. S., Barsukov, I. L. & Bystrov, V. F. (1985) *FEBS Lett.* **180**, 33–39.
- Tian, F. & Cross, T. A. (1999) *J. Mol. Biol.* **5**, 1993–2003.

## Structural Consequences of Anesthetic and Nonimmobilizer Interaction with Gramicidin A Channels

Pei Tang,\*<sup>#</sup> Virgil Simplaceanu,<sup>§</sup> and Yan Xu\*<sup>#</sup>

\*Department of Anesthesiology and Critical Care Medicine and <sup>#</sup>Department of Pharmacology, University of Pittsburgh; and <sup>§</sup>Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania 15261 USA

**ABSTRACT** Although interactions of general anesthetics with soluble proteins have been studied, the specific interactions with membrane bound-proteins that characterize general anesthesia are largely unknown. The structural modulations of anesthetic interactions with synaptic ion channels have not been elucidated. Using gramicidin A as a simplified model for transmembrane ion channels, we have recently demonstrated that a pair of structurally similar volatile anesthetic and nonimmobilizer, 1-chloro-1,2,2-trifluorocyclobutane (F3) and 1,2-dichlorohexafluorocyclobutane (F6), respectively, have distinctly different effects on the channel function. Using high-resolution NMR structural analysis, we show here that neither F3 nor F6 at pharmacologically relevant concentrations can significantly affect the secondary structure of the gramicidin A channel. Although both the anesthetic F3 and the nonimmobilizer F6 can perturb residues at the middle section of the channel deep inside the hydrophobic region in the sodium dodecyl sulfate micelles, only F3, but not F6, can significantly alter the chemical shifts of the tryptophan indole N-H protons near the channel entrances. The results are consistent with the notion that anesthetics cause functional change of the channel by interacting with the amphipathic domains at the peptide-lipid-water interface.

### INTRODUCTION

The molecular targets for general anesthetic action have proved peculiarly difficult to determine. A superfamily of ligand-gated synaptic ion channels, including the  $\gamma$ -aminobutyric acid<sub>A</sub> (GABA<sub>A</sub>) receptor, glycine receptor, neuronal nicotinic acetylcholine receptor, and 5-hydroxytryptamine<sub>3</sub> receptor, has been considered the top candidates because of their supersensitivity to general anesthetics. Recent studies (Forman et al., 1995; Mihic et al., 1997) showed that a simple substitution of a single amino acid in some of these ligand-gated ion channels can greatly change the sensitivity to general anesthetics. Although sensitivity alone cannot serve as a criterion for unequivocal identification of the sites of action, these mutagenesis findings nevertheless support the idea that general anesthetics exert their primary action by interacting with proteins (Franks and Lieb, 1994). It remains unclear, however, whether these residues constitute part of the anesthetic-binding sites, or they are involved only in allosteric linkage (Franks and Lieb, 1997). A specific structural requirement for anesthetic binding on membrane proteins has not been elucidated (Eckenhoff and Johansson, 1997).

Three-dimensional (3D) structural analysis is not yet possible for the authentic ligand-gated ion channels because of their size and structural complexity. We recently showed (Xu et al., 1998) that gramicidin A (HCO-L-Val<sup>1</sup>-Gly<sup>2</sup>-L-Ala<sup>3</sup>-D-Leu<sup>4</sup>-L-Ala<sup>5</sup>-D-Val<sup>6</sup>-L-Val<sup>7</sup>-D-Val<sup>8</sup>-L-Trp<sup>9</sup>-D-Leu<sup>10</sup>-

L-Trp<sup>11</sup>-D-Leu<sup>12</sup>-L-Trp<sup>13</sup>-D-Leu<sup>14</sup>-L-Trp<sup>15</sup>-NHCH<sub>2</sub>CH<sub>2</sub>OH), a simple cation channel with well-resolved 3D structure (Arseniev et al., 1985; Lomize et al., 1992; Cross, 1997), can serve as a model for the study of interaction of general anesthetics with transmembrane proteins. We showed that a volatile anesthetic, 1-chloro-1,2,2-trifluorocyclobutane (F3), interacted specifically with the tryptophan residues of gramicidin A near the channel entrances, whereas a structurally similar nonimmobilizer (nonanesthetic), 1,2-dichlorohexafluorocyclobutane (F6), had no specific interaction with these regions. The direct functional consequence of this was that F3 could increase the unidirectional rates of Na<sup>+</sup> transport across the gramicidin A channel, whereas F6 had no effects on Na<sup>+</sup> transport.

In the present study, we use high-resolution NMR spectroscopy to investigate possible structural changes in the gramicidin A channel after interaction with F3 or F6 takes place. We show that although neither F3 nor F6 at pharmacological concentrations can produce measurable changes in the secondary structure of the gramicidin A channel, F3, but not F6, can significantly alter the tryptophan side-chain association with the interfacial water or with the lipid headgroup.

### MATERIALS AND METHODS

#### Materials

Purified gramicidin A was purchased from Calbiochem (La Jolla, CA). Deuterated sodium dodecyl sulfate (SDS-d<sub>22</sub>) and D<sub>2</sub>O were obtained from Cambridge Isotope Laboratories (Andover, MA). F3 and F6 were purchased from PCR Inc. (Gainesville, FL). Other chemicals, of analytical grade, were from Sigma Co. (St. Louis, MO). SDS was recrystallized in ethanol before use. All other compounds were used without further purification.

Received for publication 21 September 1998 and in final form 17 February 1999.

Address reprint requests to Dr. Pei Tang, W-1357 Biomedical Science Tower, University of Pittsburgh, Pittsburgh, PA 15261. Tel.: 412-383-9798; Fax: 412-648-9587; E-mail: tang@smtp.anes.upmc.edu.

© 1999 by the Biophysical Society

0006-3495/99/05/2346/05 \$2.00

## Sample preparation

To determine anesthetic and nonanesthetic effects on channel conformation, it is critically important to minimize the amount of organic solvents in the peptide samples, for many of the solvents are general anesthetics themselves. To achieve high NMR spectral resolution in the liquid state, gramicidin A was incorporated in SDS micelles rather than in lipid bilayers. The structure of gramicidin A in the channel conformation is known to be very similar in these two environments (Cross, 1994, 1997; Weinstein et al., 1979; Killian et al., 1994; Ketchum et al., 1997; Mobashery et al., 1997). To prepare gramicidin A channel in SDS micelles, the procedure developed by Killian et al. (1994) was modified and used. Briefly, a 25 mM solution of gramicidin A in 2,2,2-trifluoroethanol (TFE) and 1000 mM SDS in H<sub>2</sub>O were prepared separately. Aliquots of gramicidin solution were added to SDS solution to reach a gramicidin-to-SDS molar ratio of 1:200. Water was then added to yield a water-to-TFE ratio of 16:1 by volume. The samples were mixed vigorously for 5 s, rapidly frozen in CO<sub>2</sub>/acetone, and lyophilized overnight at -50°C. The lyophilized samples were further vacuumed for at least 24 h to ensure nearly complete removal of TFE. The amount of TFE remaining in the samples was less than 100  $\mu$ M, as determined by GC in selected samples and confirmed by the nonexistence of any <sup>19</sup>F resonance in <sup>19</sup>F-NMR spectra before the addition of fluorinated anesthetics or nonimmobilizers. For NMR measurement, the dry samples were rehydrated with deionized water (90% H<sub>2</sub>O and 10% D<sub>2</sub>O for field-lock purposes). In each NMR sample, the gramicidin A concentration ranged from 1.9 to 2.5 mM, the pH was adjusted to 4.8, and the solution volume was 0.5 ml in a 5-mm high-precision NMR tube, which was later sealed, leaving a 2.0-ml vapor space above the solution.

F3 or F6 was titrated directly into the samples in the NMR tube with a Hamilton microsyringe. After equilibrating with the vapor phase, the total F3 or F6 concentrations in the SDS solution were estimated by <sup>19</sup>F NMR, with reference to an external standard of 0.19 mM trifluoroacetic acid (TFA) in a 10 mm NMR tube, which was coaxial with the 5-mm sample tube.

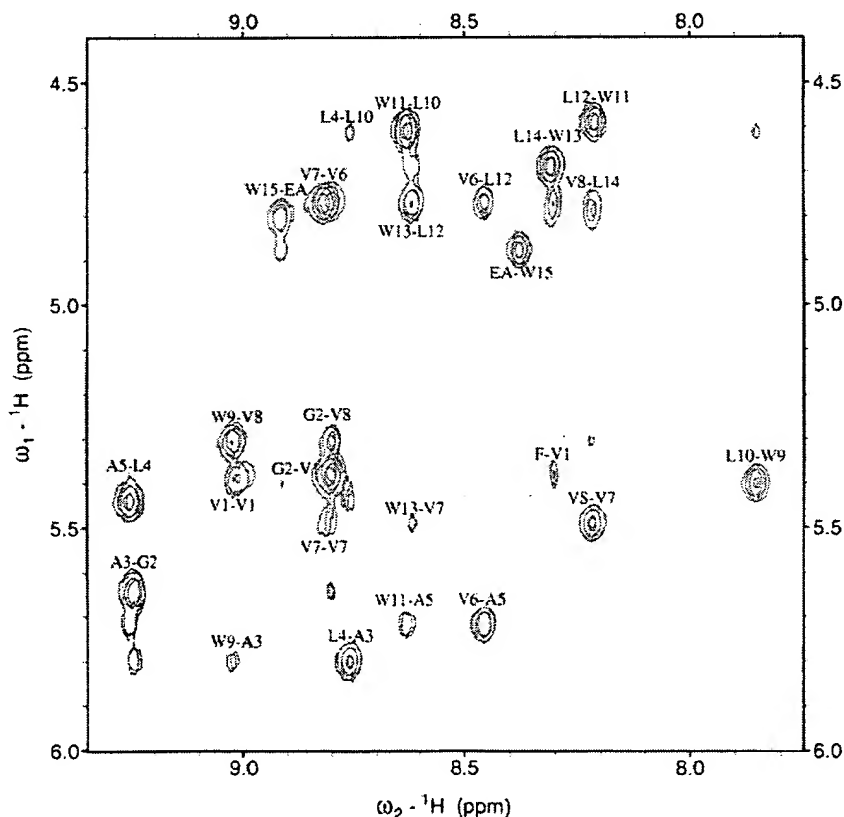
## NMR spectroscopy

High-resolution <sup>1</sup>H NMR spectra of the rehydrated micelles containing gramicidin A were recorded on Bruker 600 and 750 spectrometers with DMX consoles, operating at the <sup>1</sup>H resonance frequencies of 600.33 and 750.13 MHz, respectively. The sample temperature was maintained at 30°C. Typical experimental parameters were 10–17- $\mu$ s 90° pulses, 1.5-s repetition delays, a 9-kHz spectral width, and WATERGATE for water suppression. For one-dimensional spectra, 64 scans were accumulated in 8192 complex points. The data were zero-filled once before Fourier transformation. For NOESY experiments, spectra were acquired using a mixing time of 100 ms, 64 averages for each  $t_1$  value after two dummy scans, a datum set of 4096 complex points with 512  $t_1$  increments, and the time proportional phase incrementation (TPPI) or States method for quadrature detection in the  $t_1$  dimension. The 2D NMR spectra were processed using the NMRPipe program developed at the National Institutes of Health. The 2D peak intensities were calculated by volume integration, using the Sparky program from University of California at San Francisco.

## RESULTS

At pharmacologically relevant concentrations, neither F3 nor F6 significantly altered the secondary structure of the gramicidin A channel. Fig. 1 shows an overlay of the fingerprint region of NOESY spectra before and after addition of 14.8 mM F3 to a sample containing 1.9 mM gramicidin A in SDS micelles. Similar results were obtained for F6. Resonance assignments of the spectra were performed based on the NOE connectivity and by comparison with the literature (Lomize et al., 1992; Arseniev et al., 1985). Except for peak Val<sup>7</sup>-Val<sup>6</sup> (V7-V6), which showed a 0.017-

FIGURE 1 Overlay of the fingerprint region of two 750-MHz <sup>1</sup>H NOESY spectra, acquired at 30°C before (green) and after (red) addition of 14.8 mM F3 to 1.9 mM gramicidin A in 380 mM SDS micelles. Cross-peaks are labeled as "amide- $\alpha$  proton," using the one-letter notation for amino acids and the sequence number in the primary structure. The mixing time was 100 ms, and the experiment time needed to acquire each NOESY spectrum was 18.5 h. Except for V7-V6, no significant changes in chemical shifts and cross-peak intensities were found in this region.



ppm shift in the Val<sup>7</sup> amide proton resonance, no significant changes in chemical shift or cross-peak intensity were found in this region. However, the resonance of all indole N-H protons in the four tryptophan side chains were significantly shifted by F3 in a concentration-dependent manner. As shown in Fig. 2, all shifts are in the up-field direction. In particular, the Trp<sup>9</sup> indole N-H proton, which is located farthest from the surface, showed the largest shift. Fig. 3 depicts the chemical shift changes in Trp<sup>9</sup> indole N-H resonance as a function of F3 or F6 concentration. Clearly, F6 in the similar concentration range showed much less perturbation to the chemical shifts in this region.

## DISCUSSION

In the channel conformation, gramicidin A forms head-to-head  $\beta^{6.3}$  helical dimers (Arseniev et al., 1985). The 3D structures of this channel are well documented from high-resolution solution-state (Arseniev et al., 1985) and solid-state (Cross, 1997) NMR. Based on the known structures of the channel, the changes in chemical shift found in this study can be interpreted by considering changes in the hydrogen bonding between the observed protons and their environments. The backbone amide proton of Val<sup>7</sup> is oriented toward the middle section of the channel (i.e., deep in the tail region of the micelle) to form a hydrogen bond with the N-terminal carbonyl group. Thus the major changes in the Val<sup>7</sup> amide proton resonance are likely caused by the F3 or F6 perturbation to this hydrogen bond. The perturbation can be specific through direct interaction of F3 or F6 with the peptide in this region, or nonspecific through possible changes in the micelle shape or diameter, which in turn place strain on the hydrogen bonding. Based on the up-field

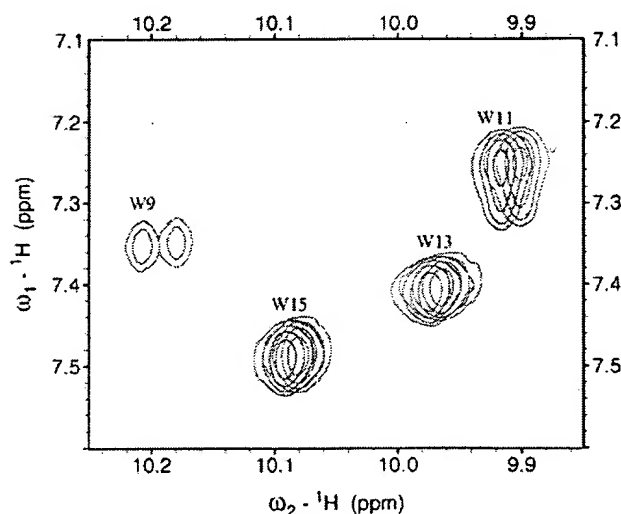


FIGURE 2 Overlay of the indole N-H region of two 750-MHz <sup>1</sup>H NOESY spectra, acquired at 30°C before (green) and after (red) addition of 14.8 mM F3 to 1.9 mM gramicidin A in 380 mM SDS micelles. The experiment time for each NOESY spectrum was 18.5 h. All resonance peaks shifted to lower frequencies; Trp<sup>9</sup> was most sensitive to F3.

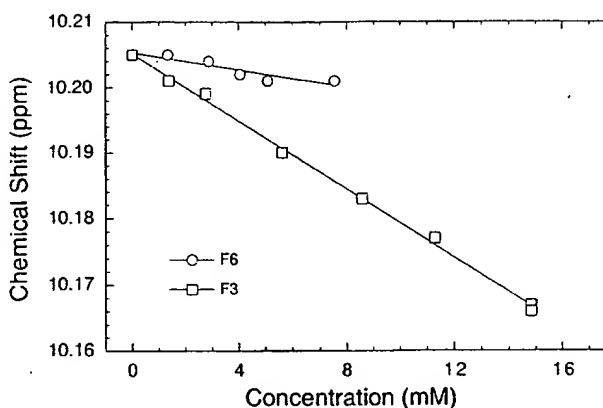


FIGURE 3 Changes in Trp<sup>9</sup> indole N-H chemical shift are plotted as a function of F3 or F6 concentration in 380 mM SDS micelles. The chemical shifts of indole N-H protons are more sensitive to F3 than to F6.

direction of the shift, it is believed that the perturbation weakens the hydrogen bonding in this region (Wagner et al., 1983). Earlier studies by others have shown that a large number of anesthetics containing the so-called acidic hydrogens have hydrogen bond-breaking effects (for a review, see Urry and Sandorfy, 1991). It has also been suggested that a good relationship may exist between hydrogen bond-breaking ability and the potency of halogenated anesthetics (Trudeau et al., 1978). Our result with F6 indicates that the fluorinated nonimmobilizer seems to have a similar ability to perturb the hydrogen bonding near the tail region of the micelles. This perturbation is in the same direction as that caused by the anesthetic F3. Thus destabilization of the dimer state by weakening of hydrogen bonding in the deep tail region of the micelle, or in the core of the lipid by analogy, seems unlikely to represent an action that is of importance to general anesthesia.

The different effects of F3 and F6 on the tryptophan side chains, however, may reveal some important characteristics associated with anesthetic modulation of transmembrane channel peptide. It has been suggested (Hu et al., 1993; Hu and Cross, 1995) that the tryptophan side chains play a critical role in anchoring the channel in the lipid membrane. The indole rings are oriented in a unique way that favors hydrogen bonding between indole N-H protons and the water molecules that either are at the surface of the membrane or penetrate into the interfacial region (Hu et al., 1993; Woolf and Roux, 1997). From the direction of changes in chemical shifts of the indole amide proton, it appears that the anesthetic F3 facilitates indole-water interaction. This is shown most profoundly for the Trp<sup>9</sup> indole N-H proton, which is farthest (~4 Å) from the surface of the micelles. It is conceivable that the amphipathic property of the anesthetic may help to reduce the energy barrier to the interaction of the Trp<sup>9</sup> side chain with the micelle-water interface. This can be achieved either by weakening any possible hydrogen bonding of Trp<sup>9</sup> indole N-H with micelle headgroups or by mediating more interfacial water mole-

cules into the Trp<sup>9</sup> indole N-H location. Hydrogen bonding of indole N-H protons with water has been shown to stabilize the cation binding at the channel entrance (Hu and Cross, 1995), a critical step in cation transport across gramicidin A channel. Indeed, our studies of Na<sup>+</sup> transport in large unilamellar vesicles showed that F3, but not F6, can significantly increase ( $p < 0.001$ ) the unidirectional rates of Na<sup>+</sup> transport across the gramicidin A channel (Xu et al., 1998). Using intermolecular truncated driven nuclear Overhauser effects (TNOE), we also confirmed that F3 did interact specifically with the tryptophan side chains. F6, in contrast, showed no measured TNOE build-up with the indole N-H protons.

The anesthetic and nonimmobilizer effects on channel dynamics may also account for some of the chemical shift changes observed. Although no attempts were made in this study to quantify the fluctuations in the channel structure, it is conceivable that by facilitating the interaction with water at the interface, where the channel is anchored, F3 may affect the channel function by altering the motion of the channel. Further studies aimed at characterizing the channel dynamics will certainly help to address this issue.

The concentrations used in this study are within the pharmacological range. We have found that the partition coefficient of F3 in SDS solution versus gas increases with increasing SDS concentration (unpublished results). At 380 mM SDS, the SDS<sub>380</sub>/gas partition coefficient at 37°C is ~13.4. Because the saline/gas partition coefficient of F3 at 37°C is 1.56 (Kendig et al., 1994), it can be estimated that the hypothetical SDS<sub>380</sub>/saline partition coefficient would be 8.6. Thus the highest F3 concentration used in this study (i.e., 14.8 mM in 380 mM SDS solution) is equivalent to ~1.7 mM in saline, which is comparable to the minimum alveolar concentration (1.47 mM in saline at 27°C) of the agent (Kendig et al., 1994).

The secondary structure of the channel is not significantly affected by either the anesthetic or the nonimmobilizer. This conclusion is true only at the anesthetic or nonimmobilizer concentrations studied. At higher concentrations, anesthetics and nonimmobilizers may exert solvent effects on the peptide, which can possibly alter the secondary structure of the channel. Moreover, gramicidin A consists of alternating L- and D-amino acids in its sequence, with the polar peptide groups lining the pore of the channel and the nonpolar side chains projecting from the exterior surface. Such an arrangement is unlikely to be found in neuronal receptor channels. Therefore, our conclusion does not rule out the possibility that structural changes may be involved in the action of general anesthetics on neuronal receptors.

It is interesting to note that in the ligand-gated ion channels, the anesthetic-sensitive sites identified by the point-mutation experiments are either within the aqueous pore (Forman et al., 1995) or at interfacial locations near the extracellular regions of the transmembrane domains on the channels (Mihic et al., 1997). At first glance, these results are rather unexpected, given the excellent correlation between the potency of general anesthetics and their solubility

in olive oil (the Meyer-Overton rule). However, as others and we have shown recently, the difference between anesthetics and nonimmobilizers lies in their ability to distribute to regions with constant access to the aqueous phase (Tang et al., 1997; North and Cafiso, 1997). Anesthetics, but not the nonimmobilizers, have the tendency to distribute to and interact with amphipathic regions in the model membranes (Xu and Tang, 1997). Thus the ability of F3 to modulate the tryptophan side chain of gramicidin channel at the amphiphilic interfacial region, and the inability of F6 to do the same, may reflect the common characteristics of anesthetic interaction with the transmembrane channel proteins. Such characteristics may be directly related to the sensitivity of the protein to general anesthetics.

In conclusion, although F3 and F6 at pharmacologically relevant concentrations did not affect the secondary structure of the gramicidin A channel, they caused distinctly different modulations of the tryptophan side chains at the amphipathic domains near the lipid interface. This difference parallels the different functional changes in the channel caused by the same anesthetic and nonimmobilizer pair.

The authors thank Dr. W. Milo Westler for assistance with NMR data acquisition, Dr. Jian Hu for anesthetic concentration calibration, and Dr. Chien Ho and Dr. Leonard L. Firestone for encouragement and constant support.

This work was supported by grants from the National Institute of General Medical Sciences, GM49202 (YX) and GM56257 (PT). The 600 MHz NMR spectrometer was obtained through an National Institutes of Health (NIH) equipment grant (S10 RR11248-01). Several experiments were carried out at the National Magnetic Resonance Facility at Madison [operation subsidized by the NIH Biomedical Research Technology Program under grant RR02301; equipment funded by the University of Wisconsin, National Science Foundation (NSF) Academic Infrastructure Program under grant BIR-9214394, the NIH Shared Instrumentation Program under grants RR02781 and RR08438, the NIH Biomedical Research Technology Program under grant RR02301, the NSF Biological Instrumentation Program under grant DMB-8415048, and the U.S. Department of Agriculture].

## REFERENCES

- Arseniev, A. S., I. L. Barsukov, V. F. Bystrov, A. L. Lomize, and A. Yu. Ovchinnikov. 1985. <sup>1</sup>H-NMR study of gramicidin A transmembrane ion channel. Head-to-head right-handed, single-stranded helices. *FEBS Lett.* 186:168-174.
- Cross, T. A. 1994. Structural biology of peptides and proteins in synthetic membrane environments by solid-state NMR spectroscopy. *Annu. Rep. NMR Spectrosc.* 29:123-167.
- Cross, T. A. 1997. Solid-state nuclear magnetic resonance characterization of gramicidin channel structure. *Methods Enzymol.* 289:672-696.
- Eckenhoff, R. G., and J. S. Johansson. 1997. Molecular interactions between inhaled anesthetics and proteins. *Pharmacol. Rev.* 49:343-367.
- Forman, S. A., K. W. Miller, and G. Yellen. 1995. A discrete site for general anesthetics on a postsynaptic receptor. *Mol. Pharmacol.* 48:574-581.
- Franks, N. P., and W. R. Lieb. 1994. Molecular and cellular mechanisms of general anaesthesia. *Nature.* 367:607-614.
- Franks, N. P., and W. R. Lieb. 1997. Inhibitory synapses. Anaesthetics set their sites on ion channels. *Nature.* 389:334-335.
- Hu, W., and T. A. Cross. 1995. Tryptophan hydrogen bonding and electric dipole moments: functional roles in the gramicidin channel and implications for membrane proteins. *Biochemistry.* 34:14147-14155.



- Hu, W., K. C. Lee, and T. A. Cross. 1993. Tryptophans in membrane proteins: indole ring orientations and functional implications in the gramicidin channel. *Biochemistry*. 32:7035–7047.
- Kendig, J. J., A. Kodde, L. M. Gibbs, P. Ionescu, and E. I. Eger. 1994. Correlates of anesthetic properties in isolated spinal cord: cyclobutanes. *Eur. J. Pharmacol.* 264:427–436.
- Ketchum, R., B. Roux, and T. Cross. 1997. High-resolution polypeptide structure in a lamellar phase lipid environment from solid state NMR derived orientational constraints. *Structure*. 5:1655–1669.
- Killian, J. A., T. P. Trouard, D. V. Greathouse, V. Chupin, and G. Lindblom. 1994. A general method for the preparation of mixed micelles of hydrophobic peptides and sodium dodecyl sulphate. *FEBS Lett.* 348:161–165.
- Lomize, A. L., V. Orekhov, and A. S. Arsen'ev. 1992. Utochnenie prostanstvennoi struktury ionnogo kanala gramitsidina A [Refinement of the spatial structure of the gramicidin A ion channel]. *Bioorg. Khim.* 18: 182–200.
- Mihic, S. J., Q. Ye, M. J. Wick, V. V. Koltchine, M. D. Krasowski, S. E. Finn, M. P. Mascia, C. F. Valenzuela, K. K. Hanson, E. P. Greenblatt, R. A. Harris, and N. L. Harrison. 1997. Sites of alcohol and volatile anaesthetic action on GABA(A) and glycine receptors. *Nature*. 389: 385–389.
- Mobashery, N., C. Nielsen, and O. S. Andersen. 1997. The conformational preference of gramicidin channels is a function of lipid bilayer thickness. *FEBS Lett.* 412:15–20.
- North, C., and D. S. Cafiso. 1997. Contrasting membrane localization and behavior of halogenated cyclobutanes that follow or violate the Meyer-Overton hypothesis of general anesthetic potency. *Biophys. J.* 72: 1754–1761.
- Tang, P., B. Yan, and Y. Xu. 1997. Different distribution of fluorinated anesthetics and nonanesthetics in model membrane: a  $^{19}\text{F}$  NMR study. *Biophys. J.* 72:1676–1682.
- Trudeau, G., K. C. Cole, R. Massuda, and C. Sandorfy. 1978. Anesthesia and hydrogen bonding: a semi-quantitative infrared study at room temperature. *Can. J. Chem.* 56:1681–1686.
- Urry, D. W., and C. Sandorfy. 1991. Chemical modulation of transmembrane protein structure and function. In *Drug and Anesthetic Effects on Membrane Structure and Function*. R. C. Aloia, C. C. Curtain, and L. M. Gordon, editors. Wiley-Liss, New York. 91–131.
- Wagner, G., A. Pardi, and K. Wüthrich. 1983. Hydrogen bond length and  $^1\text{H}$  NMR chemical shifts in proteins. *J. Am. Chem. Soc.* 105:5948–5949.
- Weinstein, S., B. A. Wallace, E. R. Blout, J. S. Morrow, and W. Veatch. 1979. Conformation of gramicidin A channel in phospholipid vesicles: a  $^{13}\text{C}$  and  $^{19}\text{F}$  nuclear magnetic resonance study. *Proc. Natl. Acad. Sci. USA*. 76:4230–4234.
- Woolf, T. B., and B. Roux. 1997. The binding site of sodium in the gramicidin A channel: comparison of molecular dynamics with solid-state NMR data. *Biophys. J.* 72:1930–1945.
- Xu, Y., and P. Tang. 1997. Amphiphilic sites for general anesthetic action? Evidence from  $^{129}\text{Xe}$ - $\{^1\text{H}\}$  intermolecular nuclear Overhauser effects. *Biochim. Biophys. Acta*. 1323:154–162.
- Xu, Y., P. Tang, and S. Liachenko. 1998. Unifying characteristics of sites of anesthetic action revealed by combined use of anesthetics and non-anesthetics. *Toxicol. Lett.* 100–101:347–352.

## Multiple alanine replacements within $\alpha$ -helix 126-134 of T4 lysozyme have independent, additive effects on both structure and stability



X.-J. ZHANG, W.A. BAASE, AND B.W. MATTHEWS

Institute of Molecular Biology, Howard Hughes Medical Institute, and  
Department of Physics, University of Oregon, Eugene, Oregon 97403

(RECEIVED November 18, 1991; REVISED MANUSCRIPT RECEIVED January 30, 1992)

### Abstract

In a systematic attempt to identify residues important in the folding and stability of T4 lysozyme, five amino acids within  $\alpha$ -helix 126-134 were substituted by alanine, either singly or in selected combinations. Together with three alanines already present in the wild-type structure this provided a set of mutant proteins with up to eight alanines in sequence. All the variants behaved normally, suggesting that the majority of residues in the  $\alpha$ -helix are nonessential for the folding of T4 lysozyme. Of the five individual alanine substitutions it is inferred that four result in slightly increased protein stability and one, the replacement of a buried leucine with alanine, substantially decreased stability. The results support the idea that alanine is a residue of high helix propensity. The change in protein stability observed for each of the multiple mutants is approximately equal to the sum of the energies associated with each of the constituent substitutions.

All of the variants could be crystallized isomorphously with wild-type lysozyme, and, with one trivial exception, their structures were determined at high resolution. Substitution of the largely solvent-exposed residues Asp 127, Glu 128, and Val 131 with alanine caused essentially no change in structure except at the immediate site of replacement. Substitutions of the partially buried Asn 132 and the buried Leu 133 with alanine were associated with modest ( $\leq 0.4$  Å) structural adjustments. The structural changes seen in the multiple mutants were essentially a combination of those seen in the constituent single replacements. The different replacements therefore act essentially independently not only so far as changes in energy are concerned but also in their effect on structure. The destabilizing replacement Leu 133  $\rightarrow$  Ala made  $\alpha$ -helix 126-134 somewhat less regular. Incorporation of additional alanine replacements tended to make the helix more uniform. For the penta-alanine variant a distinct change occurred in a crystal-packing contact, and the "hinge-bending angle" between the amino- and carboxy-terminal domains changed by  $3.6^\circ$ . This tends to confirm that such hinge-bending in T4 lysozyme is a low-energy conformational change.

**Keywords:** alanine; lysozyme; protein folding; protein structure; thermostability

In a systematic attempt to identify residues important in the folding and stability of phage T4 lysozyme we previously substituted four alanines within the  $\alpha$ -helix that includes residues 126-134 (Zhang et al., 1991). The helix is amphipathic, located on the surface of the carboxy-terminal domain, and is remote from the active site (Fig. 1; Kinemage 1). In wild-type lysozyme this  $\alpha$ -helix already includes three alanines. It was found that substitution of three solvent-exposed residues, Glu 128, Val 131, and Asn

132, with alanine did not interfere with folding or activity, and the multiple mutant in which all three of these residues were replaced with alanine (E128A/V131A/N132A) had a melting temperature at pH 2.0 that was  $3.4^\circ\text{C}$  higher than wild type. The results therefore also supported the idea that alanine is a residue of high helix propensity (Marqusee et al., 1989; Dao-pin et al., 1990; Lyu et al., 1990; Merutka et al., 1990; O'Neil & DeGrado, 1990). It was also noted in the prior study that replacement of the buried residue Leu 133 with alanine was substantially destabilizing.

In the present report we extend these studies by including the additional substitution D127A and by con-

Reprint requests to: B.W. Matthews, Institute of Molecular Biology, Howard Hughes Medical Institute, University of Oregon, Eugene, Oregon 97403.

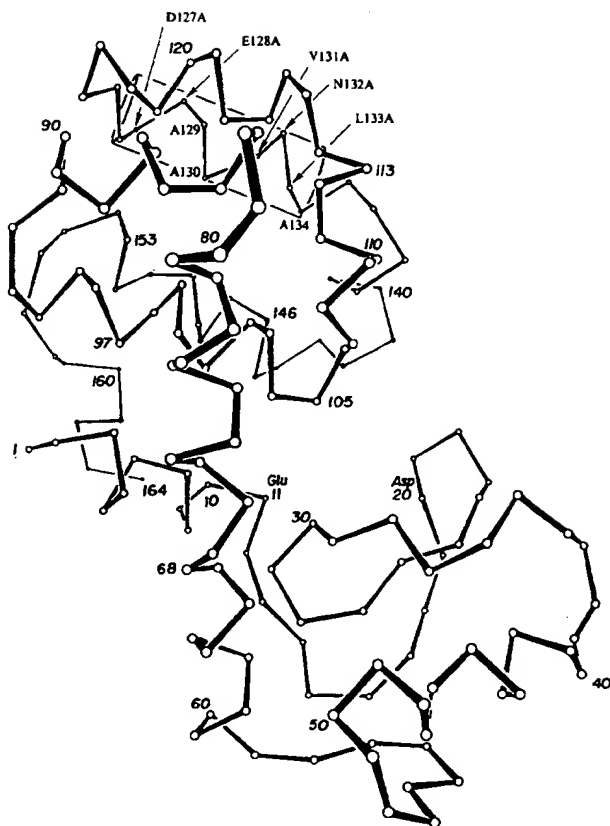


Fig. 1. Schematic drawing of the backbone structure of T4 lysozyme showing the location of the polyalanine helix and the mutations discussed in the text.

structuring additional multiple mutants in which selected alanine replacements were combined in the same protein. In total, 10 different mutants have been characterized (Table 1), culminating in the variant D127A/E128A/V131A/N132A/L133A in which five substituted alanines, together with the naturally occurring ones at positions 129, 130, and 134, result in a string of eight consecutive alanines.

## Results

The mutant lysozymes that are the subject of this study are summarized in Table 1. First there are three variants with single amino acid substitutions, Glu 128 → Ala (E128A), V131A, and L133A. These were then combined with two additional substitutions Asp 127 → Ala and Asn 132 → Ala to give the double mutants, D127A/E128A, E128A/V131A, and V131A/N132A. Further combinations provided the triple mutant E128A/V131A/N132A, which will be further abbreviated to 128/131/132, the quadruple mutants 127/128/131/132 and 128/131/132/133, and the quintuple mutant 127/128/131/132/133. The first quadruple variant has six alanines in sequence,

from residue 127 to 132, and is also referred to as A127-132, the second quadruple variant has seven alanines in sequence (A128-134), and the quintuple mutant (A127-134) has eight consecutive alanines.

Notwithstanding the introduction of up to five additional alanines, all of the mutant lysozymes behaved normally and could be purified by the standard procedure (Muchmore et al., 1989; Poteete et al., 1991).

The stabilities of the mutant lysozymes as measured by reversible unfolding at pH 2.0 are given in Table 2. The overall result is that all mutants except those that included replacement of the buried residue Leu 133 with alanine resulted in an increase in thermostability.

All of the mutants could be crystallized using conditions similar to those for wild-type lysozyme (Weaver & Matthews, 1987) and gave crystals large enough for high-resolution data collection. Prior to X-ray exposure the crystals were equilibrated with a solution of 1.05 M  $K_2HPO_4$ , 1.26 M  $NaH_2PO_4$ , 0.23 M NaCl, 1.4 mM  $\beta$ -mercaptoethanol, pH 6.7. The oscillation photographic method was used to collect X-ray data for the variants L133A, E128A/V131A, V131A/N132A, and 128/131/132. A multiwire detector (San Diego Multiwire Systems) (Xuong et al., 1985) was used for the variants D127A/E128A, 128/131/132/133, 127/128/131/132, and 127/128/131/132/133. Glu 128 of T4 lysozyme is completely solvent exposed and quite mobile (average thermal factor for atoms in the carboxyl group is  $77 \text{ \AA}^2$ ). It was therefore expected that the structure of E128A would be virtually identical with that of the wild type, and no attempt was made to solve the crystal structure of the single mutant. An early study, albeit at moderate resolution, showed that the substitution Glu 128 → Lys caused very little change in the structure of T4 lysozyme (Grütter & Matthews, 1982). Verification of the minimal structural changes associated with the substitution Glu 128 → Ala is, in any event, provided by the crystal structures of E128A/V131A and other multiple mutants in which it is incorporated.

The structural changes associated with the various mutants were first visualized using difference electron density maps and then refined (Tronrud et al., 1987) using procedures essentially as described by Dao-pin et al. (1991). The refined model of wild-type lysozyme (Bell et al., 1991) was used as the starting point for refinement. In cases where the unit cell dimensions of the mutant crystal differed significantly from those of the wild type, rigid-body refinement was first used to place the whole molecule and/or the separate amino-terminal and carboxy-terminal domains within the mutant unit cell (Dao-pin et al., 1991). Data collection and refinement statistics are summarized in Table 3. Coordinates of the refined structures have been deposited in the Brookhaven Protein Data Bank.

In the following paragraphs we will briefly describe the structural changes associated with the different variants.

Table 1. Sequences of mutant lysozymes<sup>a</sup>

Lysozyme	$\alpha$ -Helix 126-134									
	126	127	128	129	130	131	132	133	134	135
Wild type	Trp	Asp	Glu	Ala	Ala	Val	Asn	Leu	Ala	Lys
E128A			Ala	(Ala)	(Ala)				(Ala)	
V131A				(Ala)	(Ala)	Ala			(Ala)	
L133A				(Ala)	(Ala)			Ala	(Ala)	
D127A/E128A		Ala	Ala	(Ala)	(Ala)				(Ala)	
E128A/V131A			Ala	(Ala)	(Ala)	Ala			(Ala)	
V131A/N132A				(Ala)	(Ala)	Ala	Ala		(Ala)	
E128A/V131A/N132A			Ala	(Ala)	(Ala)	Ala	Ala		(Ala)	
127/128/131/132		Ala	Ala	(Ala)	(Ala)	Ala	Ala		(Ala)	
128/131/132/133			Ala	(Ala)	(Ala)	Ala	Ala	Ala	(Ala)	
127/128/131/132/133		Ala	Ala	(Ala)	(Ala)	Ala	Ala	Ala	(Ala)	
Solvent accessibility	0.27	1.10	0.76	0.00	0.14	0.78	0.34	0.01	0.37	0.85

<sup>a</sup> The locations of the alanines introduced in the different mutant lysozymes are shown. Alanines in parentheses are present in both wild-type and mutant variants. The last line gives the solvent accessibility of the side chain present in the crystal structure of wild-type lysozyme. Solvent accessibility is defined as the ratio of the area of the side chain exposed to solvent in the folded structure relative to the area exposed to solvent in an extended model peptide of the same amino acid sequence.

Table 2. Thermal stabilities of mutant lysozymes<sup>a</sup>

Mutant	$\Delta H$ (kcal/mol)	$\Delta T_m$ (°C)	$\Delta\Delta G$ (kcal/mol)
Wild type <sup>b</sup>	86.0		
E128A <sup>b</sup>	85.0	0.6 ± 0.25	0.16
V131A <sup>b</sup>	88.7	1.0 ± 0.25	0.27
L133A	53.0	-17.0 ± 2.0	-4.19
D127A/E128A	85.8	0.9 ± 0.25	0.24
E128A/V131A <sup>b</sup>	93.0	1.5 ± 0.25	0.44
V131A/N132A <sup>b</sup>	82.0	2.3 ± 0.25	0.57
E128A/V131A/N132A <sup>b</sup>	88.3	3.4 ± 0.22	0.91
127/128/131/132	86.3	4.0 ± 0.25	1.01
128/131/132/133	63.9	-10.3 ± 0.5	-2.59
127/128/131/132/133	62.2	-9.4 ± 0.5	-2.27

<sup>a</sup> All measurements are at pH 2.02, 0.2 M KCl.  $\Delta H$  is the enthalpy of unfolding at the melting temperature,  $T_m$ .  $\Delta T_m$  is the difference between the melting temperature of the mutant and that of wild-type lysozyme (40.75 °C).  $\Delta\Delta G$ , the difference between the free energy of unfolding of the mutant and wild-type proteins, was estimated using a thermodynamic model (Brandts & Hunt, 1967; Becktel & Schellman, 1987), which includes a constant change in heat capacity,  $\Delta C_p$ , estimated in this case to be 2.4 kcal/mol-deg. For the mutants whose melting temperatures are within a few degrees of wild type, the estimated error in  $\Delta\Delta G$  is ±0.1 kcal/mol. For the unstable mutants, however, especially L133A, the accuracy of  $\Delta\Delta G$  is limited by the choice of the model used for its determination and by the choice of  $\Delta C_p$ . In addition, under the conditions used in this study (0.20 M KCl, pH 2.02), which were chosen to be consistent with the prior analysis (Zhang et al., 1991), the melting of L133A above its  $T_m$  shows some departure from two-state behavior. For these reasons the uncertainty in  $\Delta\Delta G$  for L133A is difficult to estimate but could be as much as ±1 kcal/mol. Under somewhat different conditions (0.025 M potassium chloride, 0.02 M potassium phosphate, pH 3.0) (Kitamura & Sturtevant, 1989) L133A exhibits two-state melting and has an estimated  $\Delta\Delta G$  of -3.6 kcal/mol (Eriksson et al., 1992).

<sup>b</sup> Data from Zhang et al. (1991).

#### Single mutant structures: L133A

D127A and N132A were not obtained as single mutants (Table 1). As explained above, the structure of E128A was not determined. The structure of V131A has been described by Dao-pin et al. (1990). This leaves L133A.

Leucine 133 is completely inaccessible to solvent. The map showing the difference in electron density between mutant and wild type (Fig. 2A) clearly indicates the loss of the leucyl side chain and also suggests some slight adjustments in neighboring parts of the structure. The refined mutant structure (Fig. 2B) indicates that the  $\alpha$ -helical residues 109-114 move ~0.3-0.5 Å toward the space vacated by the leucyl side chain. The movement of these residues, as well as adjustments (~0.4 Å) at the substitution site itself, are also seen in a "shift plot" (Figs. 3A, 4A) showing the shift of each residue in L133A relative to wild-type lysozyme (animated in Kinemage 2). In wild-type lysozyme there is a van der Waals contact (3.6 Å) between atoms within the side chains of Leu 133 and Phe 114. In the mutant, the side chain of Phe 114 moves 0.32 Å toward the space vacated by Leu 133. Also the hydroxyl of Ser 117 moves 0.47 Å, its  $\chi^1$  angle changing from -66° to -73°. The crystallographic thermal factor of the  $\beta$ -carbon of Leu 133 increases from 13 Å<sup>2</sup> in wild type to 38 Å<sup>2</sup> in L133A, indicating substantially greater mobility in the latter structure.

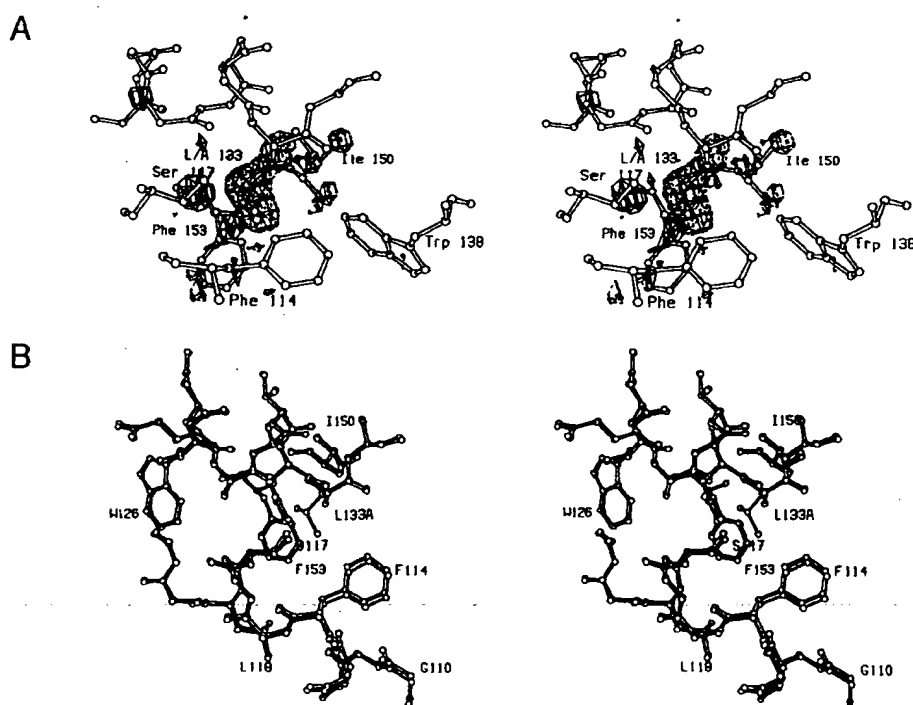
#### D127A/E128A and E128A/V131A

These two variants are considered together because Asp 127, Glu 128, and Val 131 are on the solvent-exposed side

**Table 3.** Data collection and refinement statistics<sup>a</sup>

Mutant	133	127/128	128/131	131/132	128/131/132	127/128/ 131/132	128/131/ 132/133	127/128/ 131/132/133
<b>Data collection</b>								
Method	Film	Multiwire	Film	Film	Film	Multiwire	Multiwire	Multiwire
Cell dimensions								
<i>a</i> , <i>b</i> (Å)	61.3	61.1	61.2	61.2	61.3	61.4	61.4	61.1
$\Delta a$ , <i>b</i> (Å)	0.5	0.7	0.6	0.6	0.5	0.4	0.4	0.7
<i>c</i> (Å)	96.2	95.8	95.8	96.2	96.3	95.7	96.5	93.4
$\Delta c$ (Å)	-0.6	-1.0	-1.0	-0.6	-0.5	-1.1	-0.3	-3.4
Completeness of data (%)	71	92	72	72	69	84	80	84
<i>R</i> <sub>merge</sub> (on <i>I</i> , %)	8.2	3.5	7.1	13.2	8.2	3.9	4.5	4.4
Isomorphous difference (%)	16.6	22.8	21.1	24.1	21.7	27.1	22.2	39.4
<b>Refinement</b>								
Reflections	10,953	15,363	11,915	11,083	11,942	11,361	17,451	12,441
Resolution (Å)	1.9	1.85	1.85	1.9	1.7	2.05	1.7	1.9
<i>R</i> -factor (%)	16.2	16.8	16.9	17.3	16.4	17.1	16.0	16.8
$\Delta$ <sub>bonds</sub> (Å)	0.015	0.015	0.015	0.016	0.015	0.013	0.013	0.016
$\Delta$ <sub>angles</sub> (°)	2.3	2.3	2.2	2.6	2.3	2.0	1.8	2.4
$\langle B \rangle$ (Å <sup>2</sup> )	25	26	28	33	22	34	19	26

<sup>a</sup> The cell dimensions of wild-type lysozyme are *a* = *b* = 61.2 Å, *c* = 96.8 Å.  $\Delta a$ ,  $\Delta b$ , and  $\Delta c$  are the changes in the cell dimensions of the mutant protein crystal relative to wild type. The average thermal factor  $\langle B \rangle$  for the atoms within the backbone of the refined wild-type model is 19.7 Å<sup>2</sup>. The isomorphous difference is the average difference between the observed structure amplitudes of the mutant and wild-type crystals.



**Fig. 2.** A: Map showing the difference in density between L133A and wild-type lysozyme. Amplitudes ( $F_{\text{mut,obs}} - F_{\text{WT,obs}}$ ) and phases calculated from the refined structure of wild-type lysozyme (Bell et al., 1991). Density contoured at  $3.5\sigma$ , where  $\sigma$  is the rms density throughout the unit cell. Positive contours drawn solid; negative contours broken. Resolution as in Table 3. The coordinates of wild-type lysozyme are superimposed. B: Superposition of the refined structure of L133A (open bonds) on that of wild-type lysozyme (solid bonds). In all such comparisons the two sets of coordinates were superimposed so as to minimize the rms discrepancy between the main-chain atoms in the respective carboxy-terminal domains (i.e., residues 81–160).

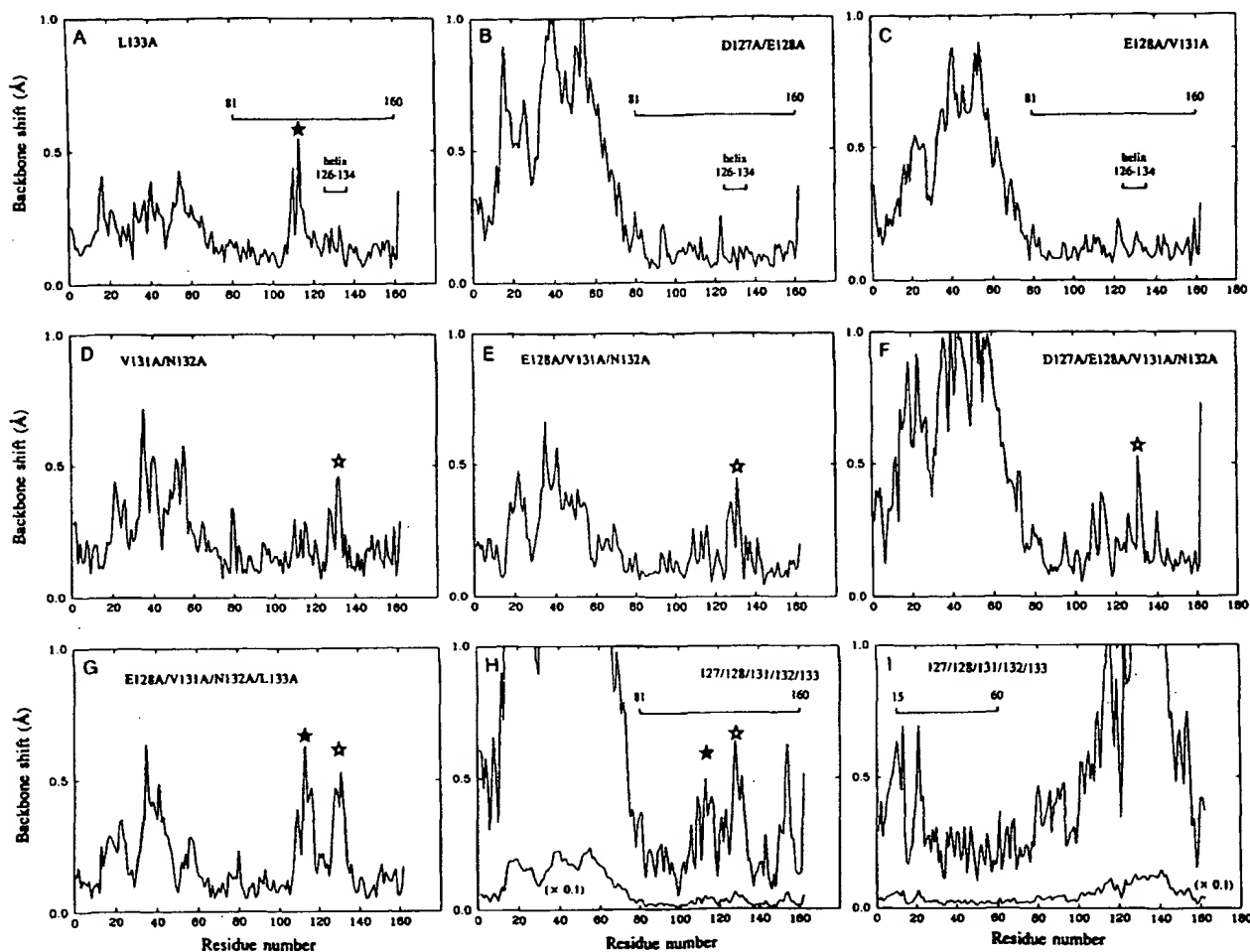


Fig. 3. Shift plots showing the displacement of the backbone atoms of each mutant relative to wild-type lysozyme. Each mutant structure was superimposed on the wild type so as to minimize the rms discrepancy between the respective backbone atoms in the carboxy-terminal domains (residues 81–160). For each amino acid the value plotted is the average (i.e., rms) discrepancy between the corresponding backbone atom ( $C^\alpha$ ,  $C^\beta$ , O, and N) in the mutant and wild-type structure. The single mutation Leu 133  $\rightarrow$  Ala causes backbone shifts in the vicinity of residues 109–114 and similar shifts are seen in all mutants that include this replacement (solid stars). Similarly, the mutation Asn 132  $\rightarrow$  Ala causes changes in the vicinity of residues 126–134 that are seen in all variants that include N132A (open stars). A: L133A; B: D127A/E128A; C: E128A/V131A; D: V131A/N132A; E: E128A/V131A/N132A; F: D127A/E128A/V131A/N132A; G: E128A/V131A/N132A/L133A; H: D127A/E128A/V131A/N132A/L133A; I: D127A/E128A/V131A/N132A/L133A; superposition based on the amino-terminal domains (residues 15–60). The large shift in Thr 21 is associated with a change in crystal contact between the backbone of Thr 21 and the side-chain of Trp 126 (see text).

of the  $\alpha$ -helix and are relatively mobile. As expected, the structural changes associated with the replacement of each of these three residues with alanine are minimal.

In the respective difference maps (Figs. 5A, 6A) there is negative density, indicating the truncation of the three side chains to alanine. In the case of Glu 128, the side-chain carboxylate in the wild-type structure is very mobile, with thermal factors above  $70 \text{ \AA}^2$ . This explains why the negative density in the difference maps does not extend to enclose the distal part of the side chain. Refinement confirms that the mutant structures are very similar to wild type (Figs. 5B, 6B). There is a slight rigid-body

movement of the amino-terminal domain relative to the carboxy-terminal domain, which can be seen in the shift plots (Figs. 3, 4; Kinemage 3). Such movements have been seen in other mutant lysozymes and are usually accompanied by a change in the  $c$  cell edge, as is the case here ( $\Delta c = -1.0 \text{ \AA}$ ). A similar, but smaller “hinge-bending” motion occurred for L133A (Fig. 3A), in which case the change in  $c$  was  $0.6 \text{ \AA}$ . Based on superposition of atoms within their carboxy-terminal domains, the root-mean-square (rms) discrepancies between D127A/E128A and wild type and E128A/V131A and wild type are  $0.11 \text{ \AA}$  and  $0.12 \text{ \AA}$ , respectively.

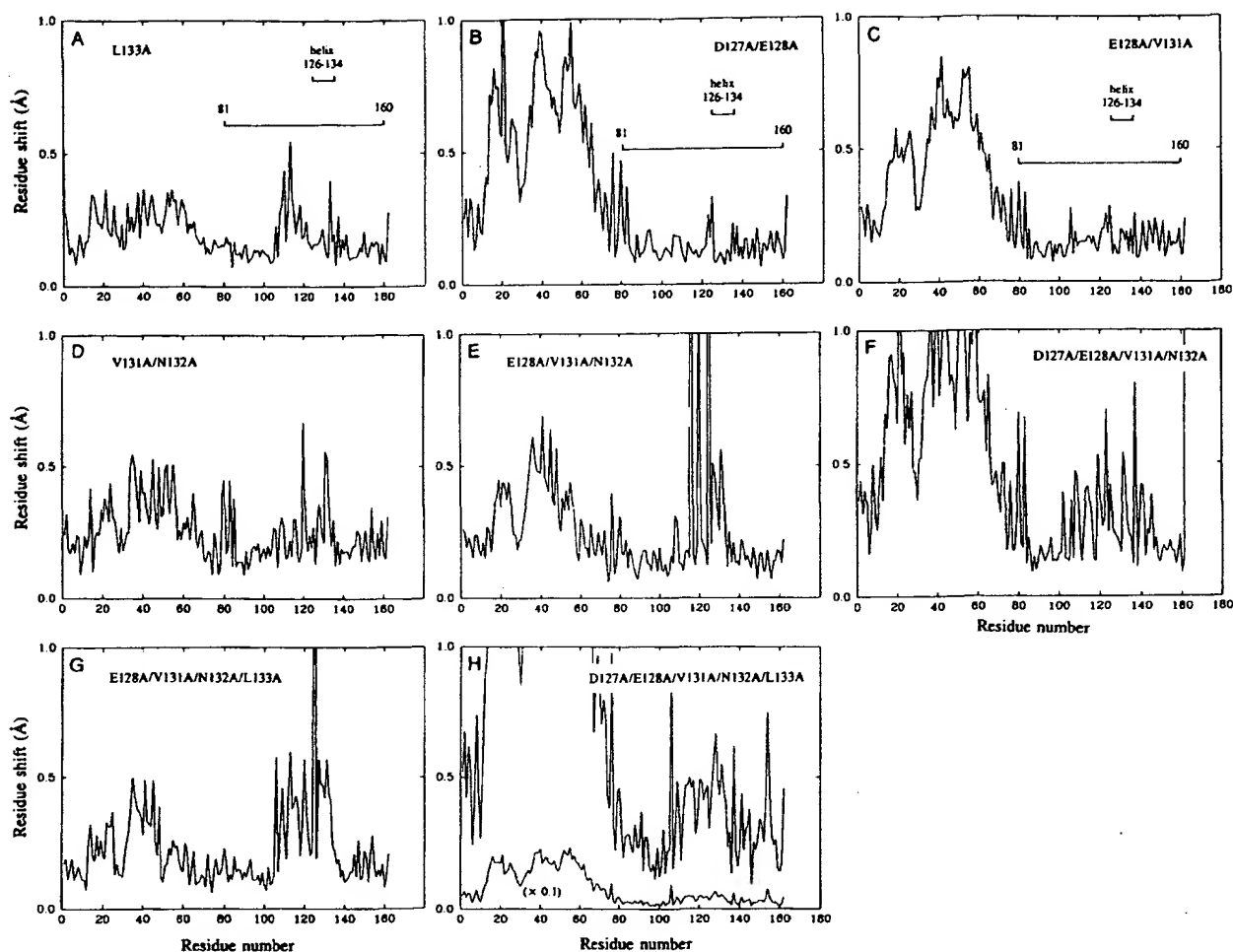


Fig. 4. Shift plots, as in Figure 3, showing the average displacement of the atoms in each residue of the mutant structure relative to wild type. A: L133A; B: D127A/E128A; C: E128A/V131A; D: V131A/N132A; E: E128A/V131A/N132A; F: D127A/E128A/V131A/N132A; G: E128A/V131A/N132A/L133A; H: D127A/E128A/V131A/N132A/L133A.

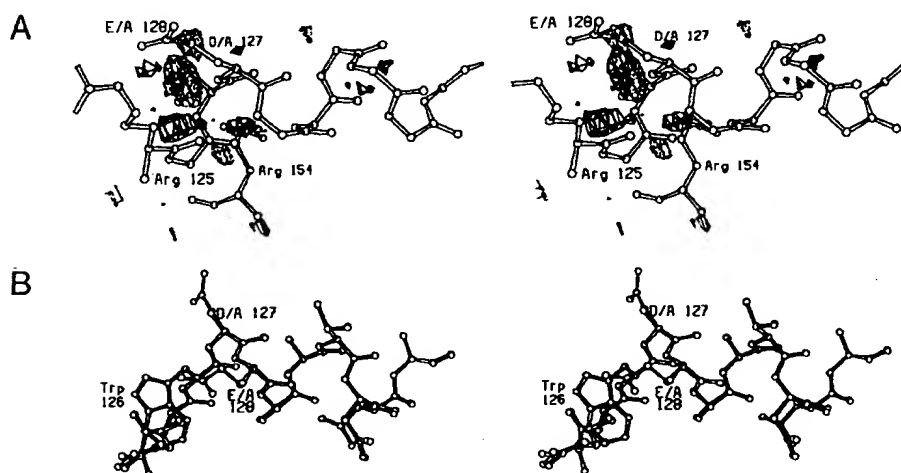


Fig. 5. A: Difference map, D127A/E128A minus wild-type lysozyme. All conventions as in Figure 2A. B: Superposition of D127A/E128A (open bonds) on wild type (solid bonds).



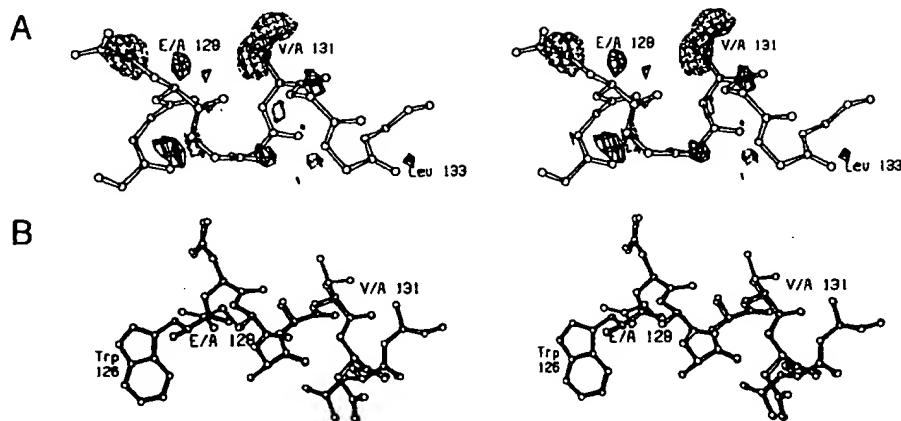


Fig. 6. A: Difference map, E128A/V131A minus wild-type lysozyme. All conventions as in Figure 2A. B: Superposition of D127A/E128A (open bonds) on wild type (solid bonds).

#### V131A/N132A and E128A/V131A/N132A

These two mutants will be considered together because they have two substitutions in common and were found

to have extremely similar structures except for the truncation of Glu 128 in the latter case.

The difference maps for the two mutants (Figs. 7A, 8A) clearly show the expected negative density corre-

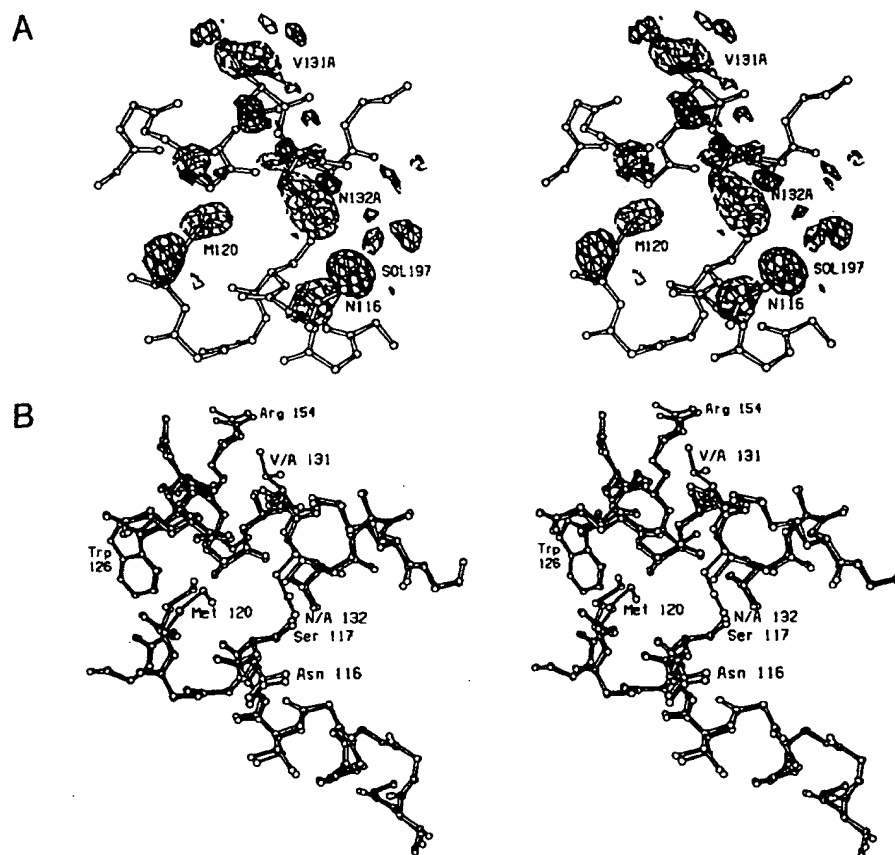


Fig. 7. A: Difference map, V131A/N132A minus wild-type lysozyme. All conventions as in Figure 2A. B: Superposition of V131A/N132A (open bonds) on wild type (solid bonds).

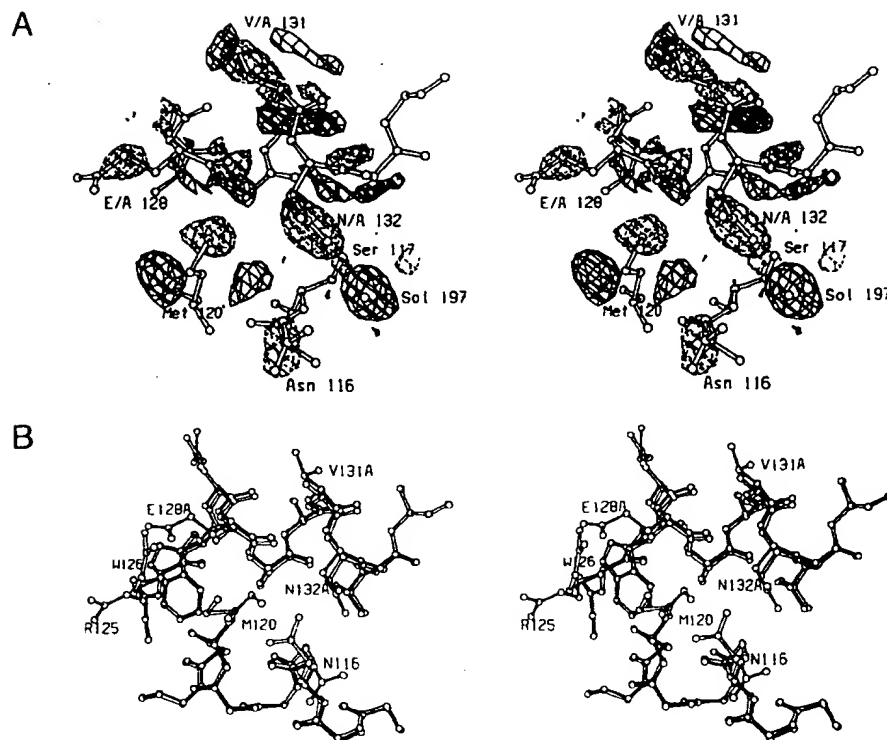


Fig. 8. A: Difference map, E128A/V131A/N132A minus wild-type lysozyme. All conventions as in Figure 2A. B: Superposition of E128A/V131A/N132A (open bonds) on wild type (solid bonds).

sponding to the substitutions of Val 131 and Asn 132 by alanine. As noted for D127A/E128A, the high mobility of Glu 128 causes the negative density at this site to be somewhat weaker. There is a strong positive density feature centered on the site occupied by solvent molecule #197 in the wild-type structure. This positive feature is a characteristic of all variants containing the Asn 132 → Ala substitution and is interpreted to be due to the replacement of solvent #197 by a chloride ion. It is presumed that the deletion of O<sup>γ1</sup> of Asn 132 favors the binding of the anion. Positive and negative density features in the vicinity of the main-chain atoms of residues 126–134 indicate a shift of the α-helix as a whole (animated in Kinemage 4). For the triple mutant, especially (Fig. 8A), positive and negative features indicate substantial (>1 Å) rearrangements in the side-chain conformations of Asn 116 and Met 120. These are seen in the refined structure of the triple mutant (Fig. 8B) and in the shift plot (Fig. 4E). In the difference map for the double mutant (Fig. 7A), there are weaker density features suggesting that Asn 116 and Met 120 may tend to undergo the same conformational adjustment as seen in the triple mutant, but the refinement (Fig. 7B) indicates that, on average, Asn 116 and Met 120 retain an essentially wild-type configuration.

The superposition of the refined structures of V131A/

N132A and E128A/V131A/N132A on wild type (Figs. 7B, 8B), as well as shift plots (Figs. 3D, 4E), shows the overall shift of the 126–134 α-helix by up to ~0.4 Å. Atoms within the 115–122 α-helix also move up to 0.25 Å. The larger movement, for helix 126–134, consists of a rotation of 3° about an axis that is approximately parallel to the axis of the 126–134 helix. The rotation is thought to be triggered by the substitution of Asn 132 by Ala, which removes a short hydrogen bond (2.5 Å) between O<sup>γ1</sup> of the asparagine and O<sup>γ</sup> of Ser 117, permitting repacking of the helix–helix interface (Zhang et al., 1991).

#### D127A/E128A/V131A/N132A

The difference map (Fig. 9A) and refined coordinates (Fig. 9B) show that the changes that occur in this mutant are similar to those seen in E128A/V131A/N132A. The α-helix undergoes a similar rotation of about 3°. A new solvent molecule occupies the site vacated by N<sup>γ2</sup> of Asn 132. Also, the nearby solvent molecule (#197) present in wild-type lysozyme is replaced by a presumed chloride ion (Fig. 9A). The side chains of Asn 116 and Met 120 appear to retain an essentially wild-type conformation, although the thermal factors of these residues increase dramatically (29 to 70 Å<sup>2</sup> and 26 to 65 Å<sup>2</sup>, respectively), indicating that the side chains are much less well ordered. It appears

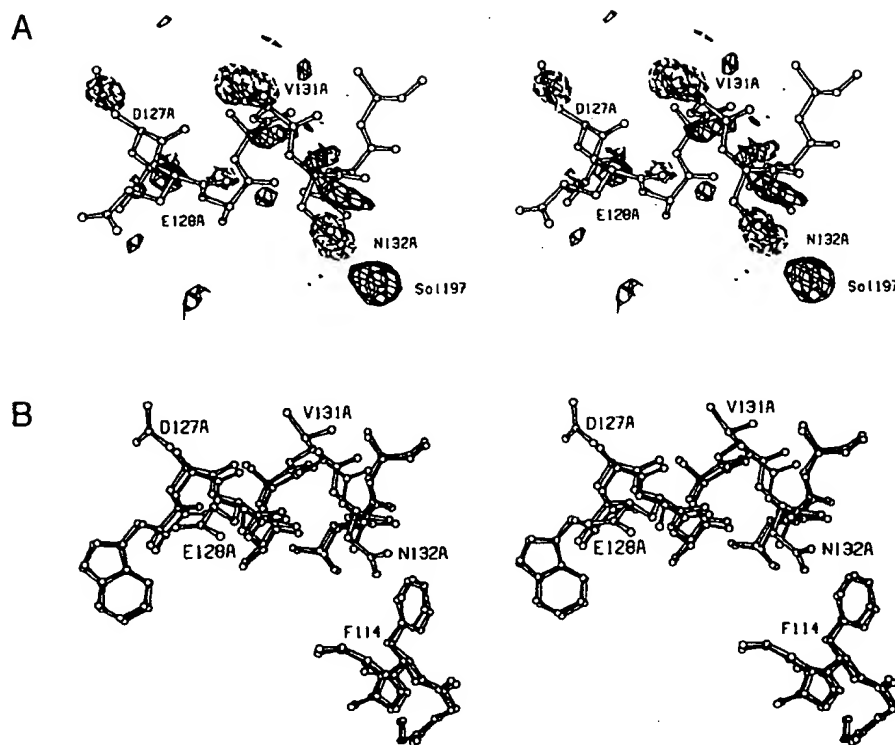


Fig. 9. A: Difference map, D127A/E128A/V131A/N132A minus wild-type lysozyme. All conventions as in Figure 2A. B: Superposition of D127A/E128A/V131A/N132A (open bonds) on wild type (solid bonds).

in this structure that both Cys 54 and Cys 97 form an adduct with  $\beta$ -mercaptoethanol (cf. Grütter et al., 1987; Bell et al., 1991). To accommodate this interaction with Cys 54, Arg 52 moves away and becomes more mobile.

#### E128A/V131A/N132A/L133A

The structural changes seen in this mutant are essentially a combination of those seen in E128A/V131A/N132A and L133A. The difference electron density map (Fig. 10A) has the expected negative density at the positions of the four truncated side chains. The other electron density features can be rationalized by the shift in the position of helix 126–134 ( $\leq 0.5$  Å) and conformational changes in the side chains of Asn 116 and Met 120 and replacement of solvent #197 by a chloride ion. The shift plot (Fig. 3A) and the superimposed structures (Fig. 10B) also show the backbone shift ( $\leq 0.5$  Å) in the vicinity of Phe 114. Also, Figure 4G shows a large apparent shift in the side chain of Arg 125. This residue is relatively mobile in both the wild-type and mutant structures.

#### D127A/E128A/V131A/N132A/L133A

One of the unusual characteristics of the crystals of this mutant is that the *c* cell dimension is 3.4 Å shorter than

that of wild-type lysozyme (Table 2). This is the largest such change observed to date in over 140 mutant lysozymes that have been crystallized isomorphously with wild type. Independent precession photographs were used to confirm that the short cell edge occurred in other crystals of this mutant and was not, for example, due to partial dehydration of the crystal used for data collection. Because of the 3% change in cell dimension, the average difference between the structure amplitudes of the mutant and wild-type lysozyme was unusually high (39%) (Table 3). For the same reason the initial difference density map (Fig. 11A) was also very noisy. Nevertheless, negative density can be seen at the locations of those residues replaced by alanine. As in all cases, refinement of the mutant structure commenced with several cycles of rigid-body refinement (Dao-pin et al., 1991). In the present instance such a procedure was practically essential.

The superposition of the mutant structure of wild type (Figs. 11B, 3H, 4H) shows that it has coordinate shifts of 0.5 Å or so through much of the carboxy-terminal domain. Most of these rearrangements were seen in the constituent mutants but, in addition, there is a structural change of about 0.6 Å at Arg 154. This change was not seen in D127/E128A or in 128/131/132/133 and seems to be a case where the combined mutant provides additional freedom for Arg 154 to move that it does not enjoy in the

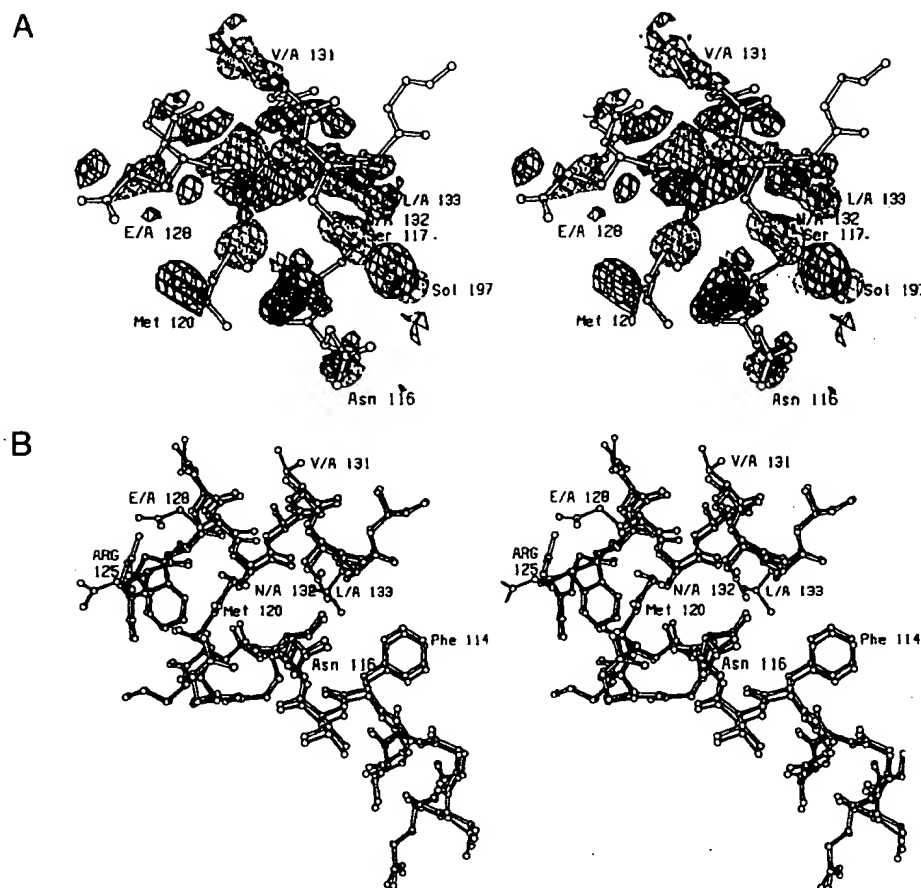


Fig. 10. A: Difference map, E128A/V131A/N132A/L133A minus wild-type lysozyme. All conventions as in Figure 2A. B: Superposition E128A/V131A/N132A/L133A (open bonds) on wild type (solid bonds). The difference map (Fig. 10A) suggests that both Met 120 and Asn 116 change their conformations in the mutant structure. The refinement indicates that these residues each are distributed between two different conformations; that shown in the figure corresponds to the conformation closest to wild-type lysozyme.

constituent mutants. It provides an example of structural changes in a mutant that are not a simple combination of those of its constituent mutations. In other words, there is an interaction between the constituent substitutions.

In the 127/128/131/132/133 mutant the accuracy of the structure in the vicinity of the cavity created by the replacement of Leu 133 with Ala is somewhat uncertain. In the final ( $2F_o - F_c$ ) electron density map the definition of the aromatic ring of Phe 153 is not perfect. Also, 2.5 Å away from the refined position of the benzyl group, in the space vacated by the Leu 133 side chain, there is an electron density feature that is of height  $5\sigma$  in the corresponding ( $F_o - F_c$ ) difference map. It seems unlikely that this density corresponds to an alternative (*t*) conformation of Phe 153 because of steric constraints. The possibility exists, therefore, that the electron density feature might indicate a water molecule in this essentially hydrophobic cavity, but we do not regard this as likely as there is no polar atom within 4 Å. A much more reliable indi-

cation is provided by the mutant structure with the single replacement Leu 133 → Ala. In this case the structure in the region of the substitution is well defined and there is no evidence whatsoever to suggest that the created cavity contains a bound solvent molecule.

## Discussion

### *Tolerance of a polyalanine helix*

The major finding of the present work is that a series of eight consecutive alanines within the amino acid sequence of T4 lysozyme does not interfere with folding or function. This supports the notion that the folding of a protein may be determined by the interactions between a subset of key amino acids (e.g., those that form the core) and that the remainder (which may constitute a relatively large fraction of the amino acid sequence) are relatively unimportant. Other evidence in support of this idea in-

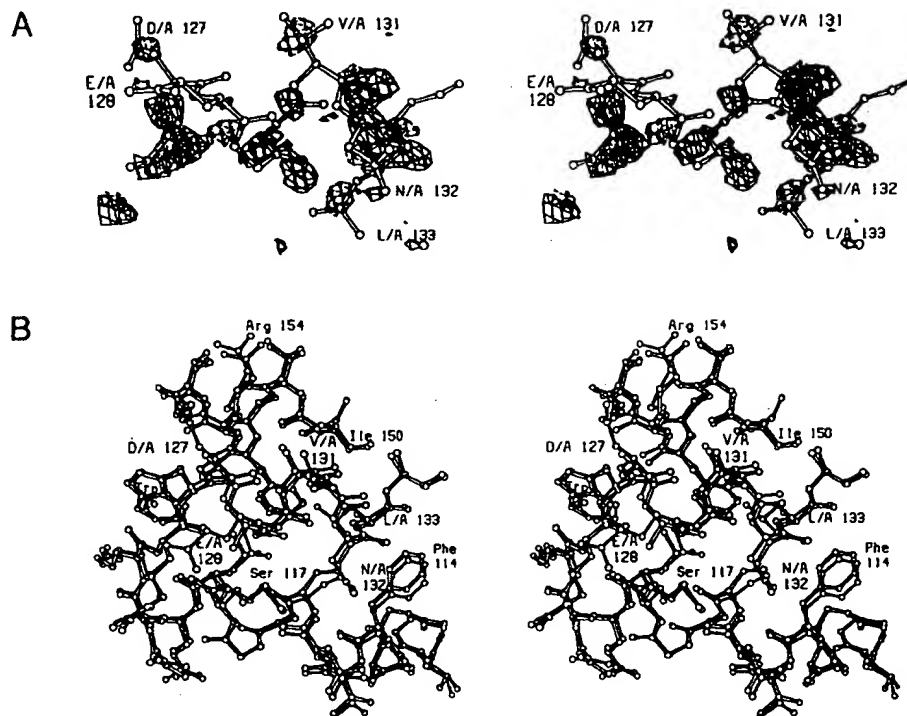


Fig. 11. A: Difference map, D127A/E128A/V131A/N132A/L133A minus wild-type lysozyme. All conventions as in Figure 2A. B: Superposition of D127A/E128A/V131A/N132A/L133A (open bonds) on wild type (solid bonds).

cludes the demonstration by Sauer and coworkers that amino acids in a given protein can have a low "information content" (Bowie et al., 1990). Also, amino acid substitutions of mobile solvent-exposed residues on the surface of a protein generally have little effect on stability (Hecht et al., 1986; Alber et al., 1987). In experiments parallel to those reported here, Heinz et al. (1992) have constructed a mutant lysozyme with 10 consecutive alanines (residues 40–49) in a different  $\alpha$ -helix and have shown that that variant also folds normally.

#### *Independence of substitutions: Energetics*

It was previously shown that substitution of alanine at the three relatively solvent-exposed sites, Glu 128, Val 131, and Asn 132, increased the thermostability of T4 lysozyme (Zhang et al., 1991). The increase in stability at the three sites was found to be approximately additive, although there was a slight synergistic effect that was interpreted to be due to an interaction between the E128A and N132A substitutions. Table 4 gives an overall summary in which an energy term is ascribed to each single alanine substitution. These energies were either measured directly or inferred by difference, as described by Zhang et al. (1991). By summing the individual energy terms and including the interaction term for those variants that include both E128A and N132A, one can predict the

change in stability expected for each of the multiple replacements. For E128A/V131A, 128/131/132, and 127/128/131/132, the agreement between the observed energy and the sum of the constituents is good. For the mutations that include L133A, the agreement is poorer, but it should be noted that L133A is a relatively unstable protein, and in such cases it is more difficult to obtain accurate values for  $\Delta\Delta G$  (Table 2). In general terms, the results support the principle of additivity and suggest that each of the substitutions acts essentially independently. Such independence would not necessarily be expected for substitutions involving pairs of amino acids that are coupled via interactions through the rest of the protein. Suppose, for example, that two amino acids in the same  $\alpha$ -helix each made contact with the core of the protein. A substitution of one amino acid might alter the alignment of the  $\alpha$ -helix relative to the rest of the protein, which, in turn, would affect the consequences of a substitution of the second amino acid. In such a case substitutions at the two sites would be coupled even though there need not be direct contact between the amino acids in question. This situation occurs with the replacements E128A and N132A, although in this case the estimated interaction energy is relatively weak ( $-0.2$  kcal/mol) (Zhang et al., 1991) (Table 4). Because of the uncertainty in the estimation of  $\Delta\Delta G$  for the destabilizing single replacement Leu 133  $\rightarrow$  Ala (Table 4), we cannot exclude

**Table 4.** Additivity of energies of stabilization<sup>a</sup>

Mutant	Stabilization observed experimentally $\Delta\Delta G$ (kcal/mol)	Stabilization inferred from constituent mutations $\Delta\Delta G$ (kcal/mol)
(D127A)		0.08
E128A	0.16	
V131A	0.27	
(N132A)		0.30
L133A	-4.19	
D127A/E128A	0.24	
E128A/V131A	0.44	0.43
V131A/N132A	0.57	
128/131/132	0.91	0.73 (0.91)
127/128/131/132	1.01	0.81 (0.99)
128/131/132/133	-2.59	-3.46 (-3.28)
127/128/131/132/133	-2.27	-3.38 (-3.20)

<sup>a</sup> The energies of stabilization of the multiple mutants are compared with the sums of energies of stabilization of the constituent replacements. The left-hand column gives the experimentally observed energies of stabilization for each of the mutant proteins ( $\Delta\Delta G$  from Table 2). Mutations D127A and N132A were not constructed as single replacements so that in these two cases the energy values are inferred by difference from the values for D127A/E128A and E128A and V131A/N132A and V131A. The sum of the energies for E128A, V131A, and N132A is 0.73 kcal/mol (as shown above), but better agreement with the experimental value is obtained by assuming that there is an interaction energy of 0.18 kcal/mol between E128A and N132A due to a structural rearrangement (see Zhang et al., 1991). This results in an energy sum of 0.91 kcal/mol, shown above in parentheses. Similarly, multiple mutants 128/131/132/133 and 127/128/131/132/133 also include both E128A and N132A, so the inferred stabilization energy is shown as the simple sum of the constituents (without parentheses) and with the 0.18-kcal/mol interaction energy term included (in parentheses).

the possibility that there is some cooperativity in the multiple mutants that include this substitution.

Table 4 suggests that the four alanine substitutions—D127A, E128A, V131A, and N132A—increase protein stability (at pH 2.0) and that the replacement of the buried Leu 133 is substantially destabilizing.

#### Independence of substitutions: Structure

An overall impression of the conformational changes associated with the different mutants is given by the shift plots shown in Figures 3 and 4 and also the comparisons of the coordinates with wild-type lysozyme (Table 5). Because the different mutations are associated with changes in the "hinge-bending angle" between the amino-terminal and carboxy-terminal domains the superposition of each structure on wild type was based on the main-chain atoms of the carboxyl domain (i.e., residues 81–160). This is the domain in which the polyalanine helix is located. Similar superpositions were also carried out for each pair of mutant structures, yielding the discrepancies shown in Table 6. Not surprisingly, mutants D127A/E128A and E128A/V131A, which involve fully solvent-exposed residues, have structures most similar to the wild type (rms

discrepancy,  $\Delta$ , of 0.11–0.12 Å) and to each other ( $\Delta = 0.13$  Å). These three structures can be considered identical, within experimental error, except in the immediate vicinity of the substitution sites (Fig. 3B,C). As discussed under Results, the replacement Asn 132 → Ala is associated with a shift of the 126–134 helix. This can be seen in all the variants that include N132A (Fig. 3D–H). Similarly, the replacement Leu 133 → Ala is associated with a shift in the  $\alpha$ -helical residues 109–114. This shift is clearly seen in all the variants that include L133A (Fig. 3A,G,H). This indicates quite clearly that the major structural changes associated with the different mutations are independent. The observation that the structural changes seen in the multiple mutants are the sum of those seen in the individual mutants is consistent with the additivity also seen in the free energies of stabilization (Table 4).

#### Large-scale changes in conformation

Superposition of the amino-terminal domain of each mutant on wild type shows that this part of the lysozyme structure is, in general, very similar in each structure (data not shown except for Fig. 3I, see below). This confirms that the overall difference between each mutant and wild type consists essentially of two parts. First there are adjustments within the carboxy-terminal domain that can include relatively extended main-chain shifts up to 0.5–0.6 Å (e.g., Fig. 3H). Second there is the rigid-body movement of the amino-terminal domain relative to the carboxyl-terminal domain, which, in the case of 127/128/131/132/133, corresponds to a rotation of 3.6° (Fig. 12). In this case, however, as can be seen in Figure 3I, there are some parts of the amino-terminal domain that do re-adjust in the mutant structure. These include residues 7–13 and 18–23. The latter is a hairpin or loop structure, which, in wild-type lysozyme, has relatively high thermal factors, indicating higher than average mobility. In the wild-type crystal lattice, there is an intermolecular hydrogen bond (3.1 Å) between the carbonyl oxygen of Thr 21 and the indole nitrogen of Trp 126 (Fig. 13; Kinemage 5). In the mutant structure this hydrogen bond is broken, and the interatomic distance increases to 4.3 Å. Instead, a new intermolecular hydrogen bond (3.1 Å) is formed between the indole nitrogen and the carbonyl oxygen of Asp 20 (Fig. 13). Associated with this change, the side chain of Thr 21 rotates from the  $g^-$  conformation in wild type to  $g^+$  in the mutant. Thus, the 3.4-Å decrease in the  $c$  cell edge is associated with hinge-bending of the two domains within the lysozyme molecule, a distinct change in a crystal-packing contact, and, as well, localized adjustments of the protein structure in the vicinity of the contact.

Although there is no direct evidence, we do not believe that the change in hinge-bending angle is an intrinsic property of the mutant lysozyme. Rather, we suggest that

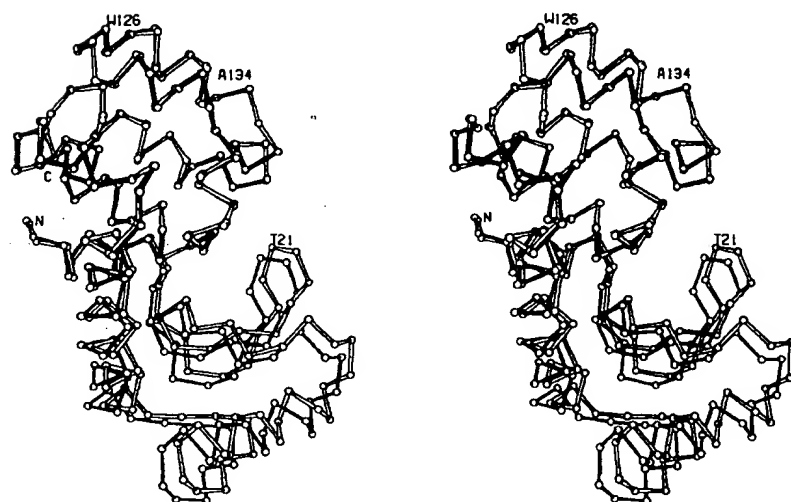
**Table 5.** Root-mean-square differences between the coordinates of polyalanine mutants and wild-type lysozyme

Mutant	Discrepancy for residues compared (Å)			
	1-162 (all atoms)	1-162 (main chain)	81-160 (all atoms)	81-160 (main chain)
133	0.18	0.16	0.18	0.16
127/128	0.26	0.22	0.16	0.11
128/131	0.22	0.19	0.17	0.12
131/132	0.25	0.21	0.24	0.18
128/131/132	0.37	0.18	0.46	0.16
127/128/131/132	0.39 <sup>a</sup>	0.28	0.29	0.19
128/131/132/133	0.29	0.20	0.35	0.22
127/128/131/132/133	0.53	0.47	0.36	0.28

<sup>a</sup> The side chain of Arg 52, which moves in association with modification of Cys 54, is not included in the comparison (see text).

**Table 6.** Root-mean-square difference between the backbone coordinates (in Å) for residues 81-160 in each of the mutant structures

	Wild type	133	127/128	128/131	131/132	128/131/132	127/128/ 131/132	128/131/ 132/133	127/128/ 131/132/133
Wild type		0.16	0.11	0.12	0.18	0.16	0.19	0.22	0.28
133			0.17	0.18	0.20	0.18	0.19	0.18	0.25
127/128				0.13	0.19	0.16	0.19	0.23	0.28
128/131					0.16	0.16	0.19	0.24	0.29
131/132						0.14	0.18	0.19	0.25
128/131/132							0.15	0.15	0.24
127/128/131/132								0.18	0.23
128/131/132/133									0.19
127/128/131/132/133									

**Fig. 12.** Stereo drawing showing the "hinge-bending" motion of mutant D127A/E128A/V131A/N132A/L133A (open bonds) relative to wild type (solid bonds). The superposition of the two structures is to optimize the agreement between their carboxy-terminal domains. The polyalanine helix includes residues 127-134. Trp 126 of one molecule and Thr 21 of another molecule participate in the contact within the crystal lattice (see text and Fig. 13).



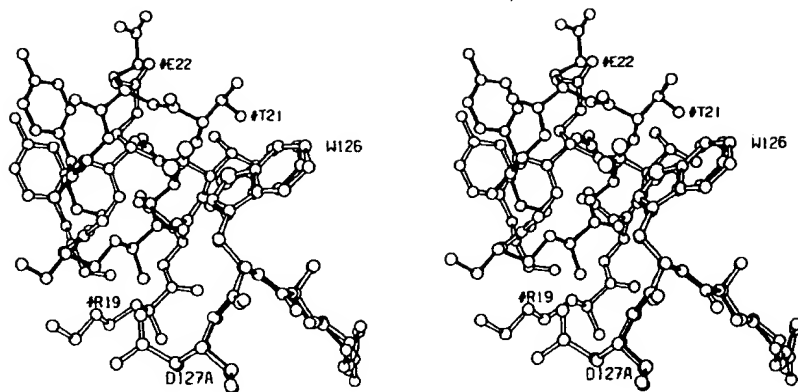


Fig. 13. Stereo drawing showing the change in crystal contact associated with the penta-alanine substitution. In wild-type lysozyme (solid bonds) there is a hydrogen bond between the indole nitrogen of Trp 126 (W126) and the carbonyl oxygen of Thr 21 in a neighboring molecule (differentiated by the # symbol) in the crystal lattice. In the D127A/E128A/V131A/N132A/L133A mutant structure (open bonds) the intermolecular hydrogen bond occurs between the indole nitrogen and the carbonyl oxygen of Asp 20. In this figure the coordinates were taken from the respective crystal structures and aligned so as to optimize the superposition of the backbone atoms of residues 125–127.

the structural changes in the vicinity of the polyalanine helix favor an alternative crystal contact, and the hinge-bending angle adjusts to facilitate this new contact. Substantial variability in hinge-bending has been observed in two other T4 lysozyme variants, Met 6 → Ile (Faber & Matthews, 1990) and Ile 3 → Pro (unpubl. results). It is presumed in the case of these variants that it requires very little energy to change the hinge-bending angle (Faber & Matthews, 1990), and the present results tend to suggest that this is true in general, as the mutations described here are well away from the hinge-bending region (Fig. 1).

#### Structure of the polyalanine helix

The polyalanine helix in the mutant 127/128/131/132/133 and in the other variants is very similar to that of wild type except for the truncation of the mutated residues. One question is whether the geometry of the  $\alpha$ -helix becomes more regular as its sequence becomes more homogeneous. In the most extreme case two full turns of the  $\alpha$ -helix consist exclusively of alanine residues. Helix geometry was investigated in several ways, first by determining the spread in the  $(\phi, \psi)$  values (Table 7). To some degree the replacement of the buried residue Leu 133 seems to make the helix less regular (i.e., increases  $\sigma(\phi)$  and  $\sigma(\psi)$ ), and the subsequent replacements tend to restore regularity, but this trend is not especially compelling. In terms of the variations of the hydrogen bond lengths within the  $\alpha$ -helix (Table 8) it also appears that the Leu 133 → Ala substitution disrupts the helix somewhat, and the subsequent substitution of additional alanines makes it more regular, in fact substantially more so than the helix in the wild-type structure. Finally, an attempt was made to compare the polyalanine helix with an ideal  $\alpha$ -helix. One of the difficulties in such a comparison is to define an ideal helix. Table 9 summarizes two different types of comparisons. In the first test, the backbone atoms (C $^\alpha$ , C, N, O) of residues 127–133 in the different mutant structures were compared with a model polyalanine  $\alpha$ -helix in which every residue has  $(\phi = -57^\circ, \psi = -47^\circ)$  (Arnott & Wonacott, 1966). In the second test, the  $\alpha$ -helix was compared with itself, but translated by a sin-

gle residue (i.e., residues 127–132 were superimposed on and compared with residues 128–133). The idea of the latter comparison was that it would help allow for a situation in which the overall helix was slightly bent. In particular, it is known (Blundell et al., 1983) that buried hydrogen bonds tend to be shorter than those exposed to solvent, and this is true for  $\alpha$ -helix 126–134 of T4 lysozyme (Table 8). The comparisons shown in Table 9 tend to support the general trend discussed above, namely that the Leu 133 → Ala replacement makes the helix less regular, and that multiple alanine replacements restore regularity. A polyalanine helix in solution might be expected to have a completely regular conformation. Here, however, not all alanines are equivalent. Some are solvent-

Table 7. Variation of  $(\phi, \psi)$  within the polyalanine helix<sup>a</sup>

Residue	$\phi$ (°)	$\psi$ (°)		
A. Ramachandran angles in wild-type lysozyme, residues 126-134				
Trp 126	-52.7	-54.8		
Asp 127	-62.6	-42.7		
Glu 128	-66.2	-39.4		
Ala 129	-64.9	-38.8		
Ala 130	-59.2	-45.0		
Val 131	-64.3	-43.5		
Asn 132	-67.4	-37.5		
Leu 133	-64.4	-29.6		
Ala 134	-72.8	-16.1		
Protein	$\langle \phi \rangle$ (°)	$\langle \psi \rangle$ (°)	$\sigma(\phi)$ (°)	$\sigma(\psi)$ (°)
B. Average $(\phi, \psi)$ values in mutant lysozymes, residues 127-133				
Wild type	-64.1	-39.5	2.5	4.8
L133A	-65.6	-37.3	4.7	8.4
D127A/E128A	-64.1	-39.8	4.8	4.6
E128A/V131A	-65.3	-38.9	6.4	3.7
V131A/N132A	-64.6	-40.3	7.7	2.8
E128A/V131A/N132A	-64.3	-39.3	4.9	3.4
127/128/131/132	-62.1	-40.8	5.8	4.9
128/131/132/133	-67.7	-37.0	4.1	4.1
127/128/131/132/133	-65.6	-39.2	3.5	4.8

<sup>a</sup>  $\langle \phi \rangle$  and  $\langle \psi \rangle$  are the rms values of  $\phi$  and  $\psi$ , and  $\sigma(\phi)$  and  $\sigma(\psi)$  are the variations for the residues within the  $\alpha$ -helix. In calculating the average values of  $\phi$  and  $\psi$  the capping residues were deleted.

**Table 8.** Helical hydrogen bond lengths within polyalanine helices

Atom pair	Bond length (Å)	
A. Hydrogen bond lengths within $\alpha$ -helix 126-134 in wild-type lysozyme		
O Trp 126 N Ala 130		3.09
O Asp 127 N Val 131		3.21
O Glu 128 N Asn 132		3.08
O Ala 129 N Leu 133		2.78
O Ala 130 N Ala 134		3.38
Protein	$\langle d \rangle$ (Å)	$\sigma(d)$ (Å)
B. Variation in hydrogen bond length in mutant structures		
Wild type	3.11	0.20
L133A	3.12	0.23
D127A/E128A	3.16	0.17
E128A/V131A	3.11	0.20
V131A/N132A	3.18	0.17
E128A/V131A/N132A	3.05	0.14
127/128/131/132	3.16	0.16
128/131/132/133	3.08	0.11
127/128/131/132/133	3.04	0.13

**Table 9.** Comparison of polyalanine helix with an ideal helix and with a "translated" helix

Protein	Agreement between residues 127–133 and an ideal helix ( $\phi = -57^\circ$ , $\psi = -47^\circ$ ) <sup>a</sup> (Å)	Agreement between residues 127–132 and residues 128–133 (Å)
Wild type	0.248	0.233
L133A	0.262	0.279
D127A/E128A	0.261	0.233
E128A/V131A	0.226	0.212
V131A/N132A	0.257	0.270
E128A/V131A/N132A	0.194	0.164
127/128/131/132	0.220	0.208
128/131/132/133	0.221	0.127
127/128/131/132/133	0.234	0.175

<sup>a</sup> If the calculation is repeated with an "average"  $\alpha$ -helix ( $\phi = -62^\circ$ ,  $\psi = -41^\circ$ ) (Barlow & Thornton, 1988), the individual discrepancies increase by about 0.04 Å, but the overall trend is the same.

exposed and some are buried. The observed conformation presumably reflects a compromise between the tendency of the repeated alanine sequence to make the helix more regular and the constraints imposed by the contacts of the  $\alpha$ -helix with the rest of the protein.

#### Surface and solvent structure

As noted above, the substitution of up to five alanines has no apparent effect on the purification of the protein or on its aggregation properties (as inferred from its behavior during purification and crystallization). Table 10

**Table 10.** Area of residues 126–134 accessible to solvent<sup>a</sup>

Mutant	Surface area for atoms specified (Å <sup>2</sup> )			
	Carbon	Oxygen	Nitrogen	Total
Wild type	218	172	45	434
I33	219	174	46	439
127/128	268	23	49	341
128/131	195	102	46	344
131/132	176	167	22	366
128/131/132	183	96	22	301
127/128/131/132	236	20	27	283
128/131/132/133	196	98	26	320
127/128/131/132/133	256	19	29	304

<sup>a</sup> Areas accessible to solvent were calculated using the method of Lee and Richards (1971) with a probe radius of  $r = 1.4$  Å.

shows the total accessible surface area of residues 126–134, given by atom type. The most significant change is in the reduction in the accessible surface area attributed to oxygen atoms due, in particular, to the loss of Asp 127 and Glu 128. Although up to five alanines are substituted, the hydrocarbon surface area exposed to solvent remains roughly constant.

Of the five mutation sites, the first three, Asp 127, Glu 128, and Val 131, are fully solvent exposed. Asn 132 is partly inaccessible to solvent, whereas Leu 133 is fully buried. For the full-exposed residues, replacement with alanine would not be expected to result in the binding of additional solvent molecules, and none is observed. Similarly, there is no evidence of solvent bound within the hydrophobic cavity created by the single Leu 133 → Ala replacement (but see the comment under Results regarding mutant 127/128/131/132/133).

The only mutants for which there is clear evidence for binding of an additional solvent molecule are those that include the replacement of Asn 132 → Ala. In E128A/V131A/N132A, 127/128/131/132, and 128/131/132/133 there is a water molecule that occupies approximately the same position as N<sup>γ2</sup> of Asn 132. In the wild-type crystal structure N<sup>γ2</sup> of Asn 132 makes a hydrogen bond (3.0 Å) to O<sup>c2</sup> of Glu 45 in a neighboring molecule. The bound solvent molecule in the mutant structures makes a similar hydrogen bond to Glu 45.

Helix 126–134 is amphiphilic. In the wild-type structure there are four water molecules that respectively hydrogen bond to the carbonyl oxygens of Asp 127, Glu 128, Ala 130, and Val 131. These hydrogen bonds have an average length of  $2.9 \pm 0.1$  Å. The average of the angles formed by the carbonyl carbon, carbonyl oxygen, and the water molecule is  $116 \pm 7^\circ$ . For Asp 127, Glu 128, and Val 131 the average of the pseudo torsion angles formed by the  $\alpha$ -carbon, carbonyl carbon, carbonyl oxygen, and water molecule is  $22 \pm 3^\circ$ . All of these values are very close to those normally observed for water molecules bound to

the solvent-exposed residues in the center of an amphiphilic  $\alpha$ -helix (Blundell et al., 1983; Baker & Hubbard, 1984; Barlow & Thornton, 1988). Because Ala 130 is partially buried it prevents the bound solvent (#216) at this position from having the standard  $22^\circ$  angle. Instead, the pseudo torsion angle is  $107^\circ$ . None of the above binding patterns of solvent is altered in any of the mutant structures (except that the thermal factors of the solvent molecules vary somewhat). In addition all of the water molecules within 3.5 Å of the side chains of residues 127, 128, 131, 132, and 133 of wild-type lysozyme retain reasonable B-values (below 65 Å<sup>2</sup>) in the mutant structures.

The overall result, therefore, is that the presence of the alanine substitutions does not alter the solvent structure on the surface of the protein except for the one case where a solvent molecule replaced a hydrogen-bonding function on a substituted amino acid.

## Methods

Methods for generation and purification of the mutant lysozymes, as well as thermodynamic and crystallographic analysis, were as described previously (Dao-pin et al., 1990; Zhang et al., 1991). The polyalanine variants were obtained by standard procedures, although in DNA sequencing it was necessary to use 7-deaza-dGTP instead of dGTP to prevent the GC-rich region (... GCX GCX GCX ...) from forming secondary structure.

## Acknowledgments

We thank Joan Wozniak and Sheila Pepiot for protein preparations, Joel Lindstrom for CD measurements, Dr. Larry Weaver for his help with the data collection, and Dr. Robert DuBose for advice on sequencing. This work was supported in part by grants from the NIH (GM21967) and the Lucille P. Markey Charitable Trust.

## References

- Alber, T., Dao-pin, S., Nye, J.A., Muchmore, D.C., & Matthews, B.W. (1987). Temperature-sensitive mutations of bacteriophage T4 lysozyme occur at sites with low mobility and low solvent accessibility in the folded protein. *Biochemistry* 26, 3754-3758.
- Arnott, S. & Wonacott, A.J. (1966). Atomic coordinates for an  $\alpha$ -helix: Refinement of the crystal structure of  $\alpha$ -poly-L-alanine. *J. Mol. Biol.* 21, 371-383.
- Baker, E.N. & Hubbard, R.E. (1984). Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* 44, 97-179.
- Barlow, D.J. & Thornton, J.M. (1988). Helix geometry in proteins. *J. Mol. Biol.* 201, 601-619.
- Becktel, W.J. & Schellman, J.A. (1987). Protein stability curves. *Biopolymers* 26, 1859-1877.
- Bell, J.A., Wilson, K.P., Zhang, X.-J., Faber, H.R., Nicholson, H., & Matthews, B.W. (1991). Comparison of the crystal structure of bacteriophage T4 lysozyme at low, medium and high ionic strengths. *Proteins Struct. Funct. Genet.* 10, 10-21.
- Blundell, T., Barlow, D., Borkakoti, N., & Thornton, J. (1983). Solvent-induced distortions and the curvature of  $\alpha$ -helices. *Nature* 306, 281-283.
- Bowie, J.U., Reidhaar-Olson, J.F., Lim, W.A., & Sauer, R.T. (1990). Deciphering the message in protein sequences: Tolerance to amino acid substitutions. *Science* 247, 1306-1310.
- Brandts, J.F. & Hunt, L. (1967). The thermodynamics of protein denaturation. III. The denaturation of ribonuclease in water and in aqueous urea and aqueous ethanol mixtures. *J. Am. Chem. Soc.* 89, 4826-4838.
- Dao-pin, S., Alber, T., Baase, W.A., Wozniak, J.A., & Matthews, B.W. (1991). Structural and thermodynamic analysis of the packing of two  $\alpha$ -helices in phage T4 lysozyme. *J. Mol. Biol.* 221, 647-667.
- Dao-pin, S., Baase, W.A., & Matthews, B.W. (1990). A mutant T4 lysozyme (Val 131  $\rightarrow$  Ala) designed to increase thermostability by the reduction of strain within an  $\alpha$ -helix. *Proteins Struct. Funct. Genet.* 7, 198-204.
- Eriksson, A.E., Baase, W.A., Zhang, X.-J., Heinz, D.W., Blaber, M., Baldwin, E.P., & Matthews, B.W. (1992). The response of a protein structure to cavity-creating mutations and its relationship to the hydrophobic effect. *Science* 255, 178-183.
- Faber, H.R. & Matthews, B.W. (1990). A mutant T4 lysozyme displays five different crystal conformations. *Nature* 348, 263-266.
- Grütter, M.G., Gray, T.M., Weaver, L.H., Alber, T., Wilson, K., & Matthews, B.W. (1987). Structural studies of mutants of the lysozyme of bacteriophage T4. The temperature sensitive mutant protein Thr 157  $\rightarrow$  Ile. *J. Mol. Biol.* 197, 315-329.
- Grütter, M.G. & Matthews, B.W. (1982). Amino acid substitutions far from the active site of bacteriophage T4 lysozyme reduce catalytic activity and suggest that the C-terminal lobe of the enzyme participates in substrate binding. *J. Mol. Biol.* 154, 525-535.
- Hecht, M.H., Sturtevant, J.M., & Sauer, R.T. (1986). Stabilization of  $\lambda$  repressor against thermal denaturation by site-directed Gly  $\rightarrow$  Ala changes in  $\alpha$ -helix 3. *Proteins Struct. Funct. Genet.* 1, 43-46.
- Heinz, D.W., Baase, W.A., & Matthews, B.W. (1992). Folding and function of a T4 lysozyme containing ten consecutive alanines illustrates the redundancy of information in an amino acid sequence. *Proc. Natl. Acad. Sci. USA* (in press).
- Kitamura, S. & Sturtevant, J.M. (1989). A scanning calorimetric study of the thermal denaturation of the lysozyme from phage T4 and the Arg 96  $\rightarrow$  His mutant form thereof. *Biochemistry* 28, 3788-3792.
- Lee, B. & Richards, F.M. (1971). The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* 55, 379-400.
- Lyu, P.C., Liff, M.I., Marky, L.A., & Kallenbach, N.R. (1990). Side-chain contributions to the stability of alpha-helical structure in peptides. *Science* 250, 669-673.
- Marqusee, S., Robbins, V.H., & Baldwin, R.L. (1989). Unusually stable helix formation in short alanine-based peptides. *Proc. Natl. Acad. Sci. USA* 86, 5286-5290.
- Merutka, G., Lipton, W., Shelonga, W., Park, S.-H., & Stellwagen, E. (1990). Effect of central-residue replacements on the helical stability of a monomeric peptide. *Biochemistry* 29, 7511-7515.
- Muchmore, D.C., McIntosh, L.P., Russell, C.B., Anderson, D.E., & Dahlquist, F.W. (1989). Expression and <sup>15</sup>N labelling of proteins for proton and nitrogen-15 NMR. *Methods Enzymol.* 177, 44-73.
- O'Neil, K.T. & DeGrado, W.F. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* 250, 646-651.
- Poteete, A.R., Dao-pin, S., Nicholson, H., & Matthews, B.W. (1991). Second-site revertants of an inactive T4 lysozyme mutant restore activity structuring the active site cleft. *Biochemistry* 30, 1425-1432.
- Tronrud, D.E., Ten Eyck, L.F., & Matthews, B.W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Crystallogr.* A43, 489-503.
- Weaver, L.H. & Matthews, B.W. (1987). Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution. *J. Mol. Biol.* 193, 189-199.
- Xuong, N.H., Nielsen, C., Hamlin, R., & Anderson, D. (1985). Strategy for data collection from protein crystals using a multiwire counter area detector diffractometer. *J. Appl. Crystallogr.* 18, 342-350.
- Zhang, X.-J., Baase, W.A., & Matthews, B.W. (1991). Toward a simplification of the protein folding problem: A stabilizing polyalanine  $\alpha$ -helix engineered in T4 lysozyme. *Biochemistry* 30, 2012-2017.

## Crystal Structure of Human Epidermal Growth Factor and Its Dimerization\*

Received for publication, April 2, 2001, and in revised form, June 27, 2001  
Published, JBC Papers in Press, July 3, 2001, DOI 10.1074/jbc.M102874200

He-Shu Lu‡, Ji-Jie Chai‡, Ming Li‡, Bing-Ren Huang§, Cun-Heng He§, and Ru-Chang Bi‡¶

From the ‡Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China and the §Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100005, China

Epidermal growth factor (EGF) is a typical growth-stimulating peptide and functions by binding to specific cell-surface receptors and inducing dimerization of the receptors. Little is known about the molecular mechanism of EGF-induced dimerization of EGF receptors. The crystal structure of human EGF has been determined at pH 8.1. There are two human EGF molecules A and B in the asymmetric unit of the crystals, which form a potential dimer. Importantly, a number of residues known to be indispensable for EGF binding to its receptor are involved in the interface between the two EGF molecules, suggesting a crucial role of EGF dimerization in the EGF-induced dimerization of receptors. In addition, the crystal structure of EGF shares the main features of the NMR structure of mouse EGF determined at pH 2.0, but structural comparisons between different models have revealed new detailed features and properties of the EGF structure.

Human epidermal growth factor (hEGF)<sup>1</sup> is a polypeptide of 53 amino acids with three internal disulfide bridges. As a mitogen, it first binds with high affinity to specific cell-surface receptors and then induces their dimerization, which is essential for activating the tyrosine kinase in the receptor cytoplasmic domain, initiating a signal transduction that results in DNA synthesis and cell proliferation (1, 2). Although EGF is a typical growth-stimulating peptide, little is known about the molecular mechanism of EGF-induced receptor dimerization. EGF was found to exist predominantly as a monomeric species in solution, and based on analyses of binding of EGF to the extracellular domain of its receptor and of the resulting dimerization of the receptor, some models have been proposed for EGF-induced dimerization of receptors (3). However, all these models have yet to be verified by structural studies, the most important of which is the structural determination of the complex of EGF with its receptor.

Considerable attention has been paid to the structural elucidation of EGF for clarification of the structure/function relationship. Although the solution structure of EGF has been

determined using NMR methods by several groups (4, 5, 6, 7, 8, 9), most of these NMR studies were performed at very acidic pH, where EGF completely loses its binding activity and biological potency (10). On the other hand, it has proved to be very difficult to grow good quality EGF crystals, despite the publication of a few relevant crystallization notes (11, 12). After numerous experiments to refine crystallization conditions, using a C-terminally truncated hEGF variant, we have obtained better quality EGF crystals in two different crystal forms at near physiological pH (13). Here we report the crystal structure of hEGF determined by the multiple isomorphous replacement method. There are two hEGF molecules in the asymmetric unit of the crystals that are in close end-to-end contact with each other. Analyses of the crystal structure and comparisons with NMR solution structures have revealed new details of the features and properties of the hEGF structure.

### EXPERIMENTAL PROCEDURES

**Materials**—The C-terminally truncated hEGF was prepared as described (14). The hEGF gene encoding 51 amino acids was chemically synthesized based on preferred code usage in yeast. The gene was under the control of the alcohol oxidase promoter and the  $\alpha$ -factor lead sequence including the 85-amino acid coding sequence. A multicopy insert was constructed as a part of the expression plasmid. The yeast *Pichia pastoris* was transformed by the expression plasmid, and a Mut<sup>+</sup> His<sup>+</sup> cell line was screened. High cell density culture of the cell line was carried out, and the cells were induced with methyl alcohol. The human epidermal growth factor with biological activity was secreted into the medium and was purified through three chromatographic steps. The final yield was 100 mg per liter of cell culture with ~98% homogeneity. All columns were purchased from Amersham Pharmacia Biotech.

**Crystallization and Data Collection**—Crystallization of hEGF was performed as described (13). The hEGF concentration was about 50 mg/ml. After successive rounds of crystallization refinements using the hanging-drop vapor-diffusion method, larger hEGF crystals were grown from a solution containing 0.9 M MgCl<sub>2</sub>, 3.5 mM CYMAL-3 (cyclohexyl-propyl-b-D-maltoside) and 0.1 M Bicine (pH 8.1) at 291 K over a period of about two months. The hEGF crystals have a typical size of 0.4 × 0.3 × 0.3 mm<sup>3</sup>, and can eventually reach a size of 0.5 × 0.5 × 0.6 mm<sup>3</sup>. These crystals belong to the space group P3<sub>1</sub>21 ( $a = b = 61.4$  Å,  $c = 87.0$  Å). They could diffract x-ray to 3.0 Å resolution at Argonne Station of synchrotron radiation (Native 1 in Table I), and to 3.2 Å on a MarResearch IP detector, using Cu K $\alpha$  x-ray from Rigaku RU-200 rotating-anode generator operating at 40 kV and 100 mA (Native 2). There are two EGF molecules (denoted by molecules A and B in the text) in the asymmetric unit of the trigonal crystals, giving a  $V_m$  of 3.82 Å<sup>3</sup>/Da (15) and a corresponding solvent content of 67.6%. The weak diffractability of hEGF crystals may be related to the higher solvent content. The difficulty in growing good quality EGF crystals may be caused by the marked conformational flexibility of the EGF structure, which will be discussed in detail in the text.

Several heavy-atom derivatives were prepared by soaking the native crystals for 3 to 7 days at 293 K in storage solution containing an appropriate concentration of dissolved heavy atom compound. Intensity data for the heavy-atom derivatives were collected at room temperature on the MarResearch IP detector, using Cu K $\alpha$  x-ray from Rigaku RU-200 rotating-anode generator operating at 40 kV and 100 mA. All

\* This work was supported by the Natural Science Foundation of China and Chinese Manned Space Engineering. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1JL9) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

¶ To whom correspondence should be addressed. Tel.: 86-(10)-64889866; Fax: 86-(10)-64871293; E-mail: rcbi@iname.com.

<sup>1</sup> The abbreviations used are: hEGF, human epidermal growth factor; RMSD, root mean-squared deviation; MIR, multiple isomorphous replacement.

TABLE I  
A summary of crystallography data

Data collection	Native1	Native2	K <sub>2</sub> PtCl <sub>6</sub>	Hg(Ac) <sub>2</sub>	UO <sub>2</sub> (NO <sub>3</sub> ) <sub>2</sub>
Data set <sup>a</sup>	P3,21	P3,21	P3,21	P3,21	P3,21
Space group	100-3.0	30-3.2	20-4.4	20-4.0	20-4.3
Resolution (Å)	$a = b = 61.43$ $c = 87.04$	$a = b = 61.21$ $c = 86.88$	$a = b = 60.86$ $c = 86.76$	$a = b = 61.17$ $c = 86.80$	$a = b = 61.03$ $c = 87.24$
Unit cell dimensions (Å)					
Number of molecules in asymmetric unit	2	2			
Observations/Unique	21097/4040	26521/3257	10710/1095	11950/1599	14001/1754
Overall completeness (%)	99.0	96.8	89.7	91.2	98.5
I/ $\sigma$ (I) (highest shell)	2.0	2.9	3.8	4.4	3.7
$R_{\text{merge}}$ (%) <sup>b</sup>	10.4	10.5	10.9	9.3	12.3
Phasing data					
$R_{\text{cullis}}$ (acentric/centric) <sup>c</sup>			0.78/0.70	0.82/0.79	0.93/0.93
Phasing power (acentric/centric) <sup>d</sup>			1.32/0.98	1.11/0.85	0.61/0.49
Occupancy			0.414	0.218	0.161
Refinement statistics					
Resolution (Å)	8.0-3.0				
R-factor <sup>e</sup>					
$R_{\text{cryst}}$ (%)	23.1				
$R_{\text{free}}$ (%)	28.3				
RMS deviations					
Bond lengths (Å)	0.007				
Bond angles (°)	1.193				

<sup>a</sup> Native1 is the data collected at Argonne Station; Native2, K<sub>2</sub>PtCl<sub>6</sub>, Hg(Ac)<sub>2</sub>, UO<sub>2</sub>(NO<sub>3</sub>)<sub>2</sub> are data collected at Marresearch Image Plate and were used in structural determination with the MIR method.

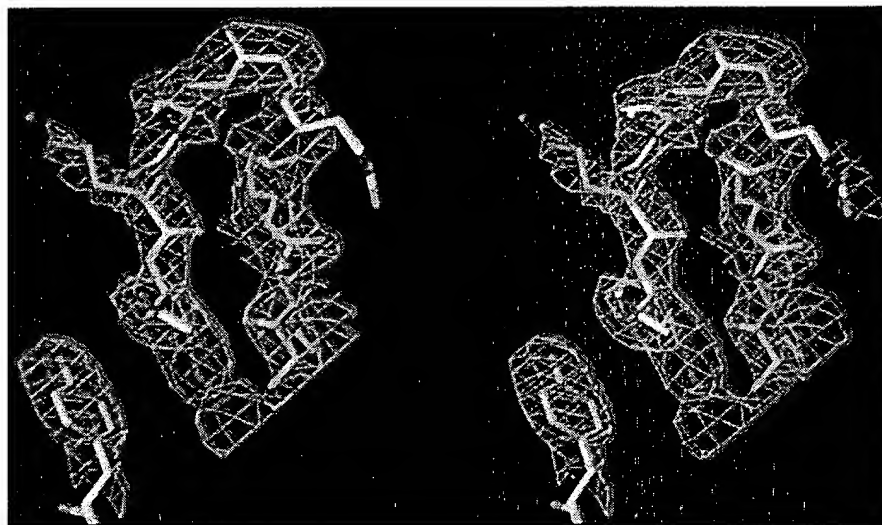
<sup>b</sup>  $R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum I_i$ , where  $I_i$  is the intensity of the  $i$ th observation and  $\langle I \rangle$  is the mean intensity of the reflection, summed over all reflections.

<sup>c</sup>  $R_{\text{cullis}} = \sum |F_{\text{PH}} - F_{\text{P}} - F_{\text{Hcalc}}| / \sum |F_{\text{PH}}|$ , where  $F_{\text{P}}$  and  $F_{\text{PH}}$  are protein and heavy-atom structure factors, respectively, and  $F_{\text{Hcalc}}$  is the calculated heavy-atom structure factor.

<sup>d</sup> Phasing power =  $\sum |F_{\text{Hcalc}}| / \sum |F_{\text{PH}}| - |F_{\text{P}}|$ .

<sup>e</sup> R-factor =  $\sum |F_{\text{o}}| - |F_{\text{c}}| / \sum |F_{\text{o}}|$ , where  $F_{\text{o}}$  and  $F_{\text{c}}$  are the observed and calculated structure factors, respectively.  $R_{\text{cryst}}$  was calculated from the 90% of reflections used in refinement, and  $R_{\text{free}}$  was calculated from the remaining 10%.

FIG. 1. A  $2F_{\text{o}} - F_{\text{c}}$  electron density map surrounding residues 21-29 and Tyr<sup>13</sup>. It is displayed in stereo using TURBO-FRODO, computed at 3.0 Å resolution and contoured at 1.5 $\sigma$ . Carbon, oxygen, and nitrogen atoms are colored yellow, red, and blue, respectively.



diffraction data were processed with the programs DENZO and SCALEPACK (16).

**Structure Determination and Refinement**—Molecular replacement studies with NMR models of mouse EGF (mEGF) and several EGF-like domains as the search model were carried out with both data sets, Native 1 and Native 2. However, all these efforts failed, perhaps due to the difference between the NMR model and the crystal structure to be discussed in the text.

The multiple isomorphous replacement (MIR) was used to determine the hEGF structure, and all calculations were performed using corresponding programs in the CCP4 package (17) and Native 2 data. The heavy atoms in the derivative crystals were found by difference Patterson method and difference Fourier maps. Heavy atom positions were refined, and phases were calculated using the program MLPHARE with reflections of  $F/\sigma(F) > 2.0$ , resulting in an initial figure of merit of 0.423 for data up to 4.0 Å. Solvent flattening and histogram matching improved the initial electron density using the program DM. It was finally determined at this point that there are two hEGF molecules in the asymmetric unit, and the boundaries between solvent and molecules

were clearly shown in the electron density map.

Based on the availability of NMR structures, the phased translation function (18) was calculated to position the hEGF molecules in the unit cell. The relevant programs in CCP4 were used with the mean NMR structure of mEGF (PDB code 1EPG, Ref. 9), as the search model. Two possible model orientations obtained from the calculation of rotation function gave clear solutions of the phased translation function. Using these solutions, the initial model of hEGF crystal structure was built with program TURBO-FRODO (19) based on the electron density and resulted in a crystallographic R-factor of 49.4%.

The refinement and rebuilding of the hEGF structure were performed, mainly using simulated annealing, conjugate gradient minimization, and group B-factor refinement protocols of the program XPLOR (20) as well as the program TURBO-FRODO. Data between 8-3.3 Å with reflections of  $F/\sigma(F) > 2.0$  were used at the early stages of refinement, later extended to 3.0 Å, and 10% data were randomly kept aside for  $R_{\text{free}}$  calculation. Fourier maps with coefficients ( $2F_{\text{o}} - F_{\text{c}}$ ) and ( $F_{\text{o}} - F_{\text{c}}$ ) were calculated in each round. In addition, simulated annealing omit maps were computed for some ambiguous regions to trace the peptide

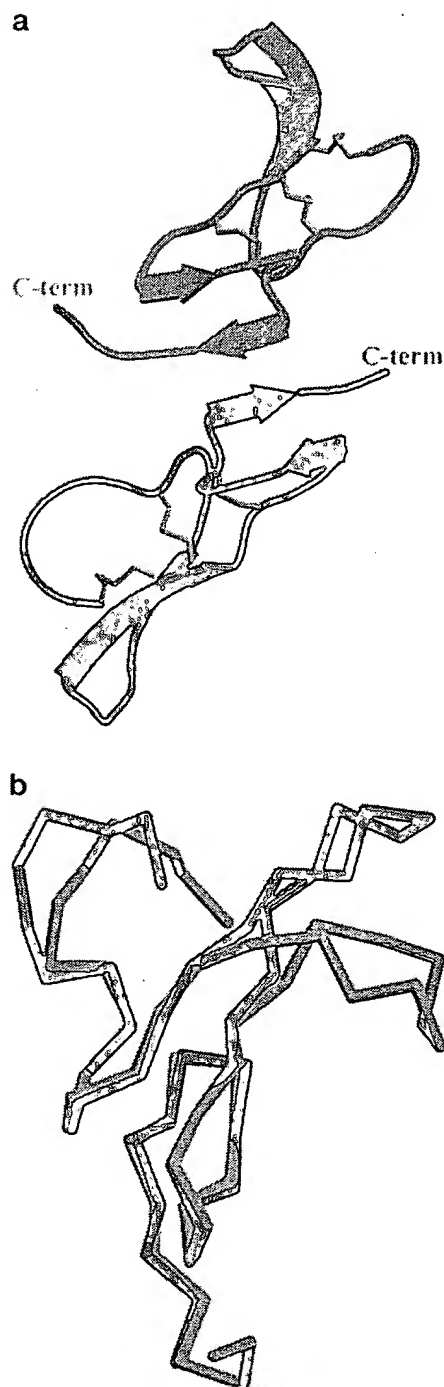


FIG. 2. The two independent hEGF molecules A (in red) and B (in green). *a*, related by a non-crystallographic 2-fold axis and form a potential dimer in the crystals. The three disulfide bridges, Cys<sup>6</sup>–Cys<sup>20</sup>, Cys<sup>14</sup>–Cys<sup>31</sup>, and Cys<sup>33</sup>–Cys<sup>42</sup> are shown in yellow. *b*, structural superposition of hEGF molecules A and B based on Cα atoms of rigid segments 13–21 and 30–47. The N-terminal segment (residues 1–12) and the residues 22–29 are adjacent to each other in the upper part of the figure. This figure was produced using MOLSCRIPT (33).

chain. During the final stages of refinement, 7 water molecules were inserted into the model. Due to the low resolution of the data, individual B-factor refinement was not carried out.

The final model was characterized by the program PROCHECK (21). Superposition between different models of the EGF structure was performed using LSQKAB in CCP4 package (17). The atomic coordinates of NMR structures of mEGF (9) were obtained from the Protein Data Bank under the accession code 1EPG and 1EPI, respectively.

## RESULTS AND DISCUSSION

**Model Quality**—A summary of the structural analysis including data collection, phasing and refinement statistics for the hEGF crystals is presented in Table I. The values of  $R_{\text{cryst}}$  and  $R_{\text{free}}$  for the refined hEGF structure are 23.1 and 28.3%, respectively for the 8.0–3.0 Å data with  $F > 2\sigma(F)$ . Amino acid residues 1–5 in both molecules A and B and the residues 48–51 in molecule A are disordered, as these regions are poorly defined in the electron density map. 98.6% of the remaining residues have appropriate backbone torsion angles in the most favorable and additionally allowed regions of the Ramachandran plot.

Except the two terminal segments, most main-chains in both hEGF molecules have well defined electron densities when contoured at  $1\sigma$  level (Fig. 1). Most side chains are also unambiguously located in the density map, whereas some polar residues on the molecular surface have poor densities showing their conformational disorder to some extent. The three disulfide bridges, Cys<sup>6</sup>–Cys<sup>20</sup>, Cys<sup>14</sup>–Cys<sup>31</sup>, and Cys<sup>33</sup>–Cys<sup>42</sup>, are also located in clearly defined electron densities.

**hEGF Structure**—The crystal structure of hEGF (Fig. 2*a*) shares main features of the NMR solution structures available. It is a structure consisting of an N-domain (residues 1–32) and a C-domain (residues 33–53). The N-domain has an irregular N-terminal peptide segment with residues 1–12 and an anti-parallel  $\beta$ -sheet (residues 19–23/28–32). The C-domain contains a short anti-parallel  $\beta$ -sheet (residues 36–38/44–46) and a C-terminal segment with residues 48–53, which are probably disordered in isolation.

Despite structural similarity between molecules A and B in the asymmetric unit, large local differences are found in peptide segments 6–12, 22–29, and 48–51, respectively (Fig. 2*b*). The most obvious difference is located at the N-terminal residues up to residue Gly<sup>12</sup>. Besides the disordered N-terminal segment with residues 1–5, residues 9–11 are well defined in molecule A, but not in molecule B. This difference may be caused by their different crystallographic environments. There are some intermolecular contacts between molecule A and a neighboring EGF molecule. They include three hydrogen bonds between residues Pro<sup>7</sup>, Ser<sup>9</sup> of molecule A, and Val<sup>35</sup>, Cys<sup>33</sup> of a neighboring molecule B in another asymmetric unit. However, there are no such kind of interactions between the N-terminal residues 6–11 of molecule B and any neighboring EGF molecules in the crystal. Another structural difference occurs at the surface turn with the residues 23–28 connecting the two anti-parallel  $\beta$ -strands of N-domain. The difference in the C-terminal peptide segment correlates with different exhibitions of electron densities for both molecules. In molecule B, residues 48–51 have clearly defined electron density, likely due to the intramolecular interactions between the residue 49 and other residues, such as the hydrogen bonds between Trp<sup>49</sup>-O and Arg<sup>45</sup>-NH1, Trp<sup>49</sup>-N and Asp<sup>46</sup>-O, whereas no electron density could be observed for residues 48–51 in molecule A.

Based on electron density map and structural comparisons between molecules A and B, several segments with rigid or flexible conformations could be defined in the EGF structure. Fig. 2*b* and Table II show that the rigid segments include residues 13–21 and 30–47. The RMSD for Cα atoms of these residues between molecules A and B is 0.517 Å. Another indication of inherent rigidity of these regions is that the average B-values of these residues are 22.14 and 26.04 Å<sup>2</sup> for main and side-chain atoms, respectively, compared with 30.07 and 36.59 Å<sup>2</sup> for the whole molecule. Two disulfide bridges, Cys<sup>14</sup>–Cys<sup>31</sup> and Cys<sup>33</sup>–Cys<sup>42</sup>, along with the highly conserved Gly<sup>18</sup> and Gly<sup>39</sup> play an important role for formation of the rigid region of the structure. The N-terminal segment (residues 1–12) and the

TABLE II  
C $\alpha$  atom RMS deviations calculated by least-square superposition between different models of the EGF structure

Models compared <sup>a</sup>	RMS Deviations (Å)				
	Residues 6–47	6–32	33–47	13–21, 30–32	13–21, 30–47
A & B	3.266	3.257	0.332	0.529	0.517 (0.252) <sup>b</sup>
A & N2	2.542	1.996	1.067	0.757	1.694 (2.751)
B & N2	3.952	3.312	1.078	0.702	1.754 (2.832)
A & N6.8	2.423	2.234	1.841	1.466	2.224 (6.497)
B & N6.8	3.769	3.235	1.799	1.558	2.313 (6.661)

<sup>a</sup> A = molecule A; B = molecule B; N6.8 = mouse NMR structure at pH 6.8; N2 = mouse NMR structure at pH 2.0.

<sup>b</sup> The values in parentheses are distances between C $\alpha$  atoms of the residue Leu<sup>47</sup>.

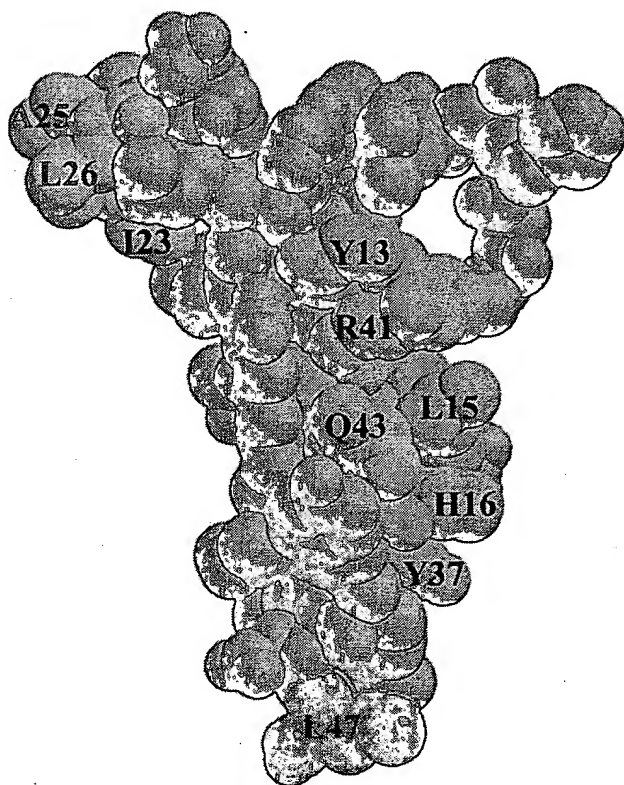


FIG. 3. Space-filling model of the EGF molecule. The figure shows the distribution of some surface residues known to be important for EGF binding to its receptor: Tyr<sup>13</sup>, Leu<sup>15</sup>, His<sup>16</sup>, Tyr<sup>37</sup>, Arg<sup>41</sup>, Gln<sup>43</sup> (in green), Ile<sup>23</sup>, Ala<sup>25</sup>, Leu<sup>26</sup> (in yellow), and Leu<sup>47</sup> (in red).

residues 23–28 concerning the irregular  $\beta$ -turn between the two  $\beta$ -strands in the major  $\beta$ -sheet, which are adjacent to each other on the head of the EGF molecule, and the C-terminal segment (residues 48–53) belong to the flexible regions of the EGF structure.

**Comparisons with NMR Structures**—Comparisons by least-square superposition based on C $\alpha$  atoms between the crystal and the NMR structures of EGF (9) were performed, and the results are shown in Table II. The whole molecule with residues 6–47 (without the disordered terminal segments), N-domain (6–32), and C-domain (33–47) were compared separately.

The comparisons of the crystal structure with the NMR structure at pH 2.0 have revealed that large structural differences are mainly located in the terminal segments, surface turns or loops, including residues 6–11, 24–27, 34–35, 39–40, and 45–47.

The flexible peptide segments (residues 1–12, 22–29, and 48–53) were excluded from more accurate comparisons between the NMR structure at pH 2.0 and the crystal structure.

The overall RMSD based on the whole molecule (residues 13–21 and 30–47) for both molecules A and B are about 1.7 Å. However, the RMSD values based on a single domain are much lower, namely, about 0.7 Å for residues 13–21 and 30–32, and about 1.07 Å for residues 33–47 (Table II). This difference may result from a relative rotation of the two domains around a hinge residue between them. The structural variation might be caused by acidic pH, thus resulting in a very different arrangement of the C-terminal segment in the EGF solution structure. Compared with the positions in the crystal structures of molecules A and B, the C $\alpha$  atom of Leu<sup>47</sup>, a crucial residue for receptor binding (22), has moved by more than 2.75 Å in the NMR structure at pH 2.0.

Regarding the comparisons with the NMR structure at pH 6.8, despite the closer pH values, the RMSD values are larger than those calculated with the NMR structure at pH 2.0 (Table II). This implies that the NMR structure at pH 6.8 may not be sufficiently accurate due to the much lower number of distance constraints used compared with calculation of the pH 2.0 NMR structure (9).

**Potential Dimer**—Molecules A and B in the asymmetric unit are in close end-to-end contact and are related by a non-crystallographic 2-fold axis. The buried area between the two molecules is  $\sim 690$  Å<sup>2</sup>, which might be enough to maintain two such small molecules together to form a dimer under certain circumstances. The dimerization interface concerns the minor  $\beta$ -sheets from the C-domains of the two molecules, forming a short four-stranded anti-parallel  $\beta$ -sheet. The non-crystallographic 2-fold axis passes the center of this four-stranded  $\beta$ -sheet (Fig. 2a).

Leu<sup>15</sup>, His<sup>16</sup>, Tyr<sup>37</sup>, Arg<sup>41</sup>, Gln<sup>43</sup>, Tyr<sup>44</sup>, Arg<sup>45</sup>, and Leu<sup>47</sup> from both hEGF molecules are involved in the intermolecular interface in the potential EGF dimer, although partial surfaces of side chains for some of these residues may be accessible. Besides the hydrogen bonds formed by atoms Gln<sup>43</sup>-O and Arg<sup>45</sup>-N in the four-stranded  $\beta$ -sheet, there are additional intermolecular hydrogen bonds between atoms His<sup>16</sup>-NE2 and Tyr<sup>37</sup>-OH in the dimer.

**Conclusions and Implications**—The crystal structure of hEGF determined at near physiological pH shares the main features of the NMR structure of mEGF determined at pH 2.0, but structural comparisons between different models revealed further details of the EGF structure. The structural differences of hEGF molecules A and B have shown detailed flexibility of the residues 22–29 in the hEGF structure. The structural comparison with the NMR structure at pH 2.0 may provide the first indication of the existence of relative movement between the N-domain and the C-domain of the EGF molecule.

The most important finding is that the dimerization of EGF molecules can occur under certain conditions. Notably, nearly all residues known to be crucial for EGF activity (23, 24), *i.e.* residues Leu<sup>15</sup>, His<sup>16</sup>, Tyr<sup>37</sup>, Arg<sup>41</sup>, Gln<sup>43</sup>, and Leu<sup>47</sup> (Fig. 3), are involved in the intermolecular interface of the potential EGF dimer. It has been reported that the EGF-receptor complex contains two EGF molecules (3). Taking into account the



importance of ligand oligomers in the ligand-induced dimerization of receptors in some cases (25, 26, 27, 28), this potential EGF dimer might be biologically relevant and play a special role in the dimerization of EGF receptors. This suggestion is inconsistent with the models proposed by Lemmon *et al.* (3) in 1997, where any EGF aggregation is excluded from the molecular details of the EGF-induced dimerization of receptors. If the hypothesis concerning EGF dimerization is correct, and the consequence of the relative movement of the N- and C-terminal domains may impede formation of the potential EGF dimer; this may then account for the inactivation of EGF at acidic pH (10).

The important residues such as Leu<sup>15</sup>, His<sup>16</sup>, Tyr<sup>37</sup>, Arg<sup>41</sup>, Gln<sup>43</sup>, and Leu<sup>47</sup> are thought to be in the site of EGF binding to its receptor (23, 24). However, according to our hypothesis they are mainly involved in the intermolecular interface of the potential EGF dimer. So there must be other receptor binding sites in the EGF molecule, which are involved in formation of a bridge to the EGF receptor. Mutation and chimera studies of EGF have indicated that some residues, *e.g.* Ile<sup>23</sup>, Ala<sup>25</sup>, Leu<sup>26</sup>, Ala<sup>30</sup>, and Asn<sup>32</sup> on the head of the EGF structure, may play an important role in the binding of EGF to its receptor (23, 29, 30). It is thought that they are mainly involved in providing a proper scaffold for the high affinity interaction between directly interacting amino acids and the receptor molecule (24, 31). As mentioned before (Fig. 2b and Table II), our structural analyses, in particular comparisons between the two molecules A and B of hEGF, have shown that the large conformational changes of these residues do not alter the type of scaffold in the EGF structure. So it is possible that the EGF head together with the variable segment containing residues 22–29 is directly involved in the interaction of EGF with its receptor. Support for this comes from a structural comparison of EGF with the 39-amino acid potato carboxypeptidase inhibitor. The latter is a low affinity EGF receptor antagonist. Its peptide segment residues 27–34 has a conformation closely similar to that of residues 22–29 in the NMR structure (32) or in the crystal structure of hEGF molecule A. The situation may resemble interferon- $\gamma$  in receptor binding, where a flexible loop of interferon- $\gamma$  is involved in the binding interface and undergoes a conformational change in the complex with the receptor (25). Therefore, important structural changes might occur in the flexible region involving residues 22–29 during EGF binding to its receptor.

In addition, if the above speculation is correct, this has implications for the heterodimerization of ErbB receptors, where there exists a 1:2 complex of ligand with receptors. It would suggest that the EGF dimerization might not play a role in the heterodimerization of ErbB receptors, and residues Leu<sup>15</sup>, His<sup>16</sup>, Tyr<sup>37</sup>, Arg<sup>41</sup>, Gln<sup>43</sup>, Leu<sup>47</sup> might play other roles in this case. Thus, further mutational studies of EGF are needed to show the different requirements for EGF-induced ErbB-1 homodimerization and for EGF-induced ErbB-1/ErbB-2 heterodimerization. However, in the process of heterodimeric formation, it could not be excluded that first a 2:2 ErbB-1 homodimer has to be formed, and subsequently ErbB-2 is involved, giving rise to a 2:4 complex, where ligand dimerization

may be also of relevance for the formation of the heterodimeric receptor complex. To verify the above hypothesis, it is crucial to determine the structure of the complex of EGF with its receptor.

**Acknowledgments**—We thank Dr. Yu Luo and Prof. Ming Luo, Department of Microbiology, Center for Biophysical Science and Engineering, University of Alabama at Birmingham for collecting the 3.0 Å resolution data at Argonne Station. We thank Prof. E. J. J. Van Zoelen, Department of Cell Biology, University of Nijmegen, The Netherlands, for his helpful discussion about the molecular mechanism of ligand-induced dimerization of receptors. We also thank Prof. Yigong Shi, Department of Molecular Biology, Princeton University, for his reading the manuscript.

#### REFERENCES

1. Carpenter, G., King, L., and Cohen, S. (1978) *Nature* **276**, 409–410
2. Carpenter, G., and Cohen, S. (1990) *J. Biol. Chem.* **265**, 7709–7712
3. Lemmon, M. A., Bu, Z., Ladbury, J. E., Zhou, M., Pinchasi, D., Lax, I., Engelman, D. M., and Schlessinger, J. (1997) *EMBO J.* **16**, 281–294
4. Cooke, R. M., Wilkinson, A. J., Baron, M., Pastore, A., Tappin, M. J., Campbell, I. D., Gregory, H., and Sheard, B. (1987) *Nature* **327**, 339–341
5. Montelione, G. T., Wüthrich, K., Nice, E. C., Burgess, A. W., and Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 5226–5230
6. Montelione, G. T., Wüthrich, K., Burgess, A. W., Nice, E. C., Wagner, G., Gibson, K. D., and Scheraga, H. A. (1992) *Biochemistry* **31**, 236–249
7. Kohda, D., Go, N., Hayashi, K., and Inagaki, F. (1988) *J. Biochem.* **103**, 741–743
8. Hommel, U., Harvey, T. S., Driscoll, P. C., and Campbell, I. D. (1992) *J. Mol. Biol.* **227**, 271–282
9. Kohda, D., and Inagaki, F. (1992) *Biochemistry* **31**, 11928–11939
10. Massagué, J. (1983) *J. Biol. Chem.* **258**, 13614–13620
11. Higuchi, Y., Morimoto, Y., Horinaka, A., and Yasuoka, N. (1988) *J. Biochem.* **103**, 905–906
12. Degenhardt, M., Weber, W., Eschenburg, S., Dierks, K., Funari, S. S., Rapp, G., and Betzel, C. (1998) *Acta Crystallogr. Series D* **54**, 999–1001
13. Chai, J. J., Li, M., Huang, B. R., Luo, Y., Luo, M., Bi, R.-C., He, C.-H. (2000) *Acta Crystallogr. Series D* **56**, 62–63
14. Huang, B. R., Cai, L. W., Liao, H. T., Tang, S. X., and Zhou, H. T. (1998) *Chinese J. Biochem. Mol. Biol.* **14**, 512–517
15. Matthews, B. W. (1968) *J. Mol. Biol.* **33**, 491–497
16. Otwinowski, Z. (1993) *Oscillation Data Reduction Program in Data Collection and Processing* (Sawyer, L., Isaacs, N., and Bailey, S., eds) pp. 56–62, Daresbury Laboratory, Warrington UK
17. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr. Series D* **50**, 760–763
18. Colman, P. M., Fehllhammer, and Bartels, K. (1976) *Crystallographic Computing Techniques*, p. 248, Munksgaard, Copenhagen
19. Jones, T. A. (1985) *Methods Enzymol.* **115**, 157–171
20. Brunger, A. T., Kuriyan, J., and Karplus, M. (1987) *Science* **235**, 458–460
21. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291
22. Matsunami, R. K., Yette, M. L., and Stevens, A. (1991) *J. Cell. Biochem.* **46**, 242–249
23. Groenen, L. C., Nice, E. C., and Burgess, A. W. (1994) *Growth factor* **11**, 235–257
24. Van Zoelen, E. J. J., Stortelers, C., Lenferink, A. E. G., and Van De Poll, M. L. M. (2000) *Vitam. Horm.* **59**, 99–131
25. Walter, M. R., Windsor, W. T., Nagabhushan, T. L., Lundell, D. J., Lunn, C. A., Zauodny, P. J., and Narula, S. K. (1995) *Nature* **376**, 230–235
26. Banner, D. W., D'Arcy, A., Janes, W., Gentz, R., Schoenfeld, H. J., Broger, C., Loetscher, H., and Lesslauer, W. (1993) *Cell* **73**, 431–445
27. Philo, J., Talvenheimo, J., Wen, J., Rosenfeld, R., Welcher, A., and Arakawa, T. (1994) *J. Biol. Chem.* **269**, 27840–27846
28. Spivak-Kroizman, T., Lemmon, M. A., Dikic, I., Ladbury, J. E., Pinchasi, D., Huang, J., Jaye, M., Crumley, G., Schlessinger, J., and Lax, I. (1994) *Cell* **79**, 1015–1024
29. Richter, A., Drummond, D. R., MacGarvie, J., Puddicombe, S. M., Chamberlin, S. G., and Davies, D. E. (1995) *J. Biol. Chem.* **270**, 1612–1616
30. Van de Poll, M. L., Van Vugt, M. J., Lenferink, A. E., and Van Zoelen, E. J. (1997) *Biochemistry* **36**, 7425–7431
31. Campbell, I. D., Cooke, R. M., Baron, M., Harvey, T. S., and Tappin, M. J. (1989) *Prog. Growth Factor Res.* **1**, 13–22
32. Blanco-Aparicio, C., Molina, M. A., Fernandez-Salas, E., Frazier, M. L., Mas, J. M., Querol, E., Aviles, F. X., and De Llorens, R. (1998) *J. Biol. Chem.* **273**, 12370–12377
33. Kraulis, P. J. (1993) *J. Appl. Crystallogr.* **26**, 283–290